

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw.model

Run on: May 29, 2003, 15:14:20 ; Search time 3987 Seconds
(without alignments)
11387.102 Million cell updates/sec

Title: US-09-934-066-1

Perfect score: 1560

Sequence: 1 ctcaagaatcagattcaa.....gaaaaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. NO. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1540	98.7	1563	8	AV120765	AV120765 Arabidops
2	1518.8	97.4	1541	8	AV084227	AV084227 Arabidops
3	832	53.3	83253	8	AF000383	AF000383 Arabidops
4	627.2	40.2	656	8	AF070375	AF070375 Arabidops
5	467.6	30.0	1576	8	AF238384	AF238384 Vigna rad
6	467.4	30.0	1980	8	BVU309173	BVU309173 Beta vulg
7	466	29.9	1659	8	D89972	D89972 Vigna mungo
8	463	29.7	1839	8	CSCYSPRNS	CSCYSPRNS
9	454.8	29.2	1485	8	AV133531	AV133531 Arabidops
10	454.8	29.2	1516	8	AV059104	AV059104 Arabidops
11	454.8	29.2	1760	8	AF424619	AF424619 Arabidops
12	454.8	29.2	1779	8	AF370160	AF370160 Arabidops
13	451.4	28.9	1850	8	PVZ99956	PVZ99956 Phaseolus v
14	444	28.5	1323	6	E05717	E05717 Asparaginil
15	442.2	28.3	1437	8	AV090296	AV090296 Arabidops
16	442.2	28.3	1616	8	AV058055	AV058055 Arabidops
17	441.6	28.3	1323	6	E05718	E05718 Asparaginil
18	435.6	27.9	1834	8	VSCYSPROB	VSCYSPROB
19	424.2	27.2	1724	8	D89971	D89971 Vigna mungo
20	421.8	27.0	1572	8	AF260827	AF260827 Ipomoea b
21	393.2	25.2	1140	8	VNA238882	VNA238882 Vicia nar
22	386	24.7	1866	8	AB025310	AB025310 Oryza sat
23	385.8	24.7	1919	8	AF082347	AF082347 Zea mays
24	385.2	24.7	1864	8	ZMAL131719	ZMAL131719 Zea mays
25	383	24.6	1529	8	LES243876	LES243876 Lycopersi
26	377.6	24.2	1762	8	RCCVPE	RCCVPE
27	376.2	24.1	1569	8	ZMAL131718	ZMAL131718 Zea mays
28	373	23.9	1840	8	VNCYSTPRO	VNCYSTPRO
29	362.6	23.2	1461	8	AV059156	AV059156 Arabidops
30	362.6	23.2	1643	8	AF367254	AF367254 Arabidops
31	361.6	23.2	1652	8	PVZ99957	PVZ99957 Phaseolus v
32	351	22.5	1736	8	SOYCPA	SOYCPA
33	347.2	22.3	1437	8	NTA238880	NTA238880 Nicotiana
34	345.6	22.2	1836	8	CENAEI	CENAEI
35	345	22.1	1323	6	E05716	E05716 Asparaginil
36	343.4	22.0	1323	6	E05720	E05720 Asparaginil
37	342	21.9	1851	8	VSA007743	VSA007743 Vicia sat
38	338.8	21.7	1715	8	AF169019	AF169019 Glycine m
39	336.4	21.6	1649	8	AF169973	AF169973 Sesamum i
40	314.4	20.2	1152	6	E05721	E05721 Asparaginil
41	306.8	19.7	1360	8	NTA238881	NTA238881 Nicotiana
42	305.4	19.6	894	6	E05722	E05722 Asparaginil
43	304	19.5	1485	8	AF082346	AF082346 Hordeum v
44	302.4	19.4	813	6	E05723	E05723 Asparaginil
45	294	18.8	1642	8	AV062178	AV062178 Oryza sat

ALIGNMENTS

RESULT 1
AV120765 1563 bp mRNA linear PLN 22-JUN-2002
LOCUS Arabidopsis thaliana vacuolar processing enzyme/asparaginyl
DEFINITION endopeptidase, putative (At3g20210) mRNA, complete cds.
ACCESSION AV120765
VERSION AV120765.1 GI:21539540
KEYWORDS FLI_CDNA.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Southwick,A., Karlin-Neumann,G., Nguyen,M., Tripp,M., Miranda,M.,

Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K., Hayashizaki, K., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (11-JUN-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PESC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Southwick, A., Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

source Location/Qualifiers

1. .1563

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAFL09-06-A19 (R15622)"

/notes="This clone is in pBluescript"

ecotype: Columbia

1. .1563

/gene="At3g20210"

/notes="MAL21.23"

63. .1463

/gene="At3g20210"

/codon_start=1

/product="vacuolar processing enzyme/asparaginyl

endopeptidase, putative"

/protein_id="AA053323.1"

/db_xref="GI:21539541"

/translation="MSPICGHQILVFLHALLIFSAESRKTOLLNDNVESDSKAG

TRAVIVASNEYNYNRHQNDICHAIQILKGLADENIIVFHDIIAFSENPRPGV

IINPDEDYKGPADYTKRAVNVQNFYNLLGNESGYVGGNKKVYKSGPNDNFIY

YADGAPLLIAMPTGDEYKADQNEVLEAKRKKYNNKVIYVERACESGMEGLIKK

NLNIYTAANSRESWGYCPESYPPPEIDTCLGDTFSLWLESDLDHDMSKETL

EQHYVYKRVGSDVPETSRVCFGRFGRMKDYLSSYIGRNPENDNFTTFESFSPIS

NSGLVNPRIPLLLYLRKIQKAPMGSLKESKAKLLDEKHKRKOIDSITDIURLSV

KQTNVLNLLSTFTTGOPLVDWDCETLVNFKNHCAGATVHYLUKYGALANICNMG

VDVQTVSAFEQACSN"

BASE COUNT 508 a 290 c 322 g 443 t

ORIGIN

Query Match 98.7%; Score 1540; DB 8; Length 1563;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1540; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCACAGAATCAGATTCAGATAGAGATTTTCAACAACATGCTAGTCTCTTGGTCAC 60

|||||

DB 24 CTCACAGAATCAGATTCAGATAGAGATTTTCAACAACATGCTAGTCTCTTGGTCAC 83

|||||

QY 61 TTTCAGATCTCTGTTTCTTCATGCTTTCATCTCTCAGCTCAGTCCCGCAAAACC 120

|||||

DB 84 TTTCAGATCTCTGTTTCTTCATGCTTTCATCTCTCAGCTCAGTCCCGCAAAACC 143

|||||

QY 121 CAATTGCTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180

|||||

DB 144 CAATTGCTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 203

|||||

181 GCTGTTTGTAGTCTGATCAAAATGAATATATTAACACAGGCATCAGGCTGACATATGC 240
 |||||
 DB 204 GCTGTTTGTAGTCTGATCAAAATGAATATATTAACACAGGCATCAGGCTGACATATGC 263
 |||||
 QY 241 CACGCGTATCAGATCTCCGAAAAGCGGTTTAAAAGATGAAAACATCATCTGTTTATG 300
 |||||
 DB 264 CACGCGTATCAGATCTCCGAAAAGCGGTTTAAAAGATGAAAACATCATCTGTTTATG 323
 |||||
 QY 301 TATGATGATATCCGTTTCTCGGAGAAATCCCTAGGCTCGAGTTATCATTAATAACCA 360
 |||||
 DB 324 TATGATGATATCCGTTTCTCGGAGAAATCCCTAGGCTCGAGTTATCATTAATAACCA 383
 |||||
 QY 361 GATGGAGAAAGATCTTATTAAGGAGTCTCTTAAGGACTACACTAAAGAGCTCTTAATGTT 420
 |||||
 DB 384 GATGGAGAAAGTGTATTAAGGAGTCTCTTAAGGACTACACTAAAGAGCTCTTAATGTT 443
 |||||
 QY 421 CAAAATCTTACAAATGTGTTTACTTGGAAATGAAAGTGGGCTCACAGGAGGAAATGGCAAA 480
 |||||
 DB 444 CAAAATCTTACAAATGTGTTTACTTGGAAATGAAAGTGGGCTCACAGGAGGAAATGGCAAA 503
 |||||
 QY 481 GTTGTGAAAAGTGGTCTCTTAATGATATATCTTTCATCTATATATGCTGAGCAGTCTCT 540
 |||||
 DB 504 GTTGTGAAAAGTGGTCTCTTAATGATATATCTTTCATCTATATATGCTGAGCAGTCTCT 563
 |||||
 QY 541 GGCTTAATACGATGCGCCACTGCTGATGAAGTTATGGCAAAAGATTTCAATCAAGTCTTTG 600
 |||||
 DB 564 GGCTTAATACGATGCGCCACTGCTGATGAAGTTATGGCAAAAGATTTCAATCAAGTCTTTG 623
 |||||
 QY 601 GAGAAGATGATGAAGAGAAAAAATACAAAGATGCTGATCTATGTTGAAGCATCTGCAAA 660
 |||||
 DB 624 GAGAAGATGATGAAGAGAAAAAATACAAAGATGCTGATCTATGTTGAAGCATCTGCAAA 683
 |||||
 QY 661 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACGACATGACTGCT 720
 |||||
 DB 684 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACGACATGACTGCT 743
 |||||
 QY 721 GCTAAATCTTAAGAGAGCAGCTGGGAGTTTACTGCTGAGTATATCTCTCTCTCTCTCT 780
 |||||
 DB 744 GCTAAATCTTAAGAGAGCAGCTGGGAGTTTACTGCTGAGTATATCTCTCTCTCTCTCT 803
 |||||
 QY 781 TCTGAGATTCGAATCTCTCGCGGATACATTTAGCATCTCTTGGCTTCAGGACAGATGAC 840
 |||||
 DB 804 TCTGAGATTCGAATCTCTCGCGGATACATTTAGCATCTCTTGGCTTCAGGACAGATGAC 863
 |||||
 QY 841 CTTTCATGATGAGCAAAAGAGACTTTTGGAGCAACAATACCACTGTTGTTAAAGAGAGATGA 900
 |||||
 DB 864 CTTTCATGATGAGCAAAAGAGACTTTTGGAGCAACAATACCACTGTTGTTAAAGAGAGATGA 923
 |||||
 QY 901 GGATCTGATGATCAGAGACTTCTCATGTATGCGGTTTCGGAAACAGAGAAAGATGCTTAAA 960
 |||||
 DB 924 GGATCTGATGATCAGAGACTTCTCATGTATGCGGTTTCGGAAACAGAGAAAGATGCTTAAA 983
 |||||
 QY 961 GATTATCTTTCTCTTTACATTTGGAAGAAATCCCTGAAAACGATAACTTCACTTTTCACGGA 1020
 |||||
 DB 984 GATTATCTTTCTCTTTACATTTGGAAGAAATCCCTGAAAACGATAACTTCACTTTTCACGGA 1043
 |||||
 QY 1021 TCCTTTTCTCTCAGCAATCTCTAAATCTTGGCTTGGTCAATCCCGGGGATATTCCTCTGCTA 1080
 |||||
 DB 1044 TCCTTTTCTCTCAGCAATCTCTAAATCTTGGCTTGGTCAATCCCGGGGATATTCCTCTGCTA 1103
 |||||
 QY 1081 TACTCTCAGAGAAAGATTTCAAAAAGCTCCCAATGGGATCACTTTGAAAGCAAAAGAGCTCAG 1140
 |||||
 DB 1104 TACTCTCAGAGAAAGATTTCAAAAAGCTCCCAATGGGATCACTTTGAAAGCAAAAGAGCTCAG 1163
 |||||
 QY 1141 AAGAAATGCTTCCAGAAAGAGATCATAGGAACAATCGATCAGAGCAATTAACAGACATT 1200
 |||||
 DB 1164 AAGAAATGCTTCCAGAAAGAGATCATAGGAACAATCGATCAGAGCAATTAACAGACATT 1223
 |||||
 QY 1201 CTCGGGCTTTCAGTTAAACCAACCAATGCTTAAATCTCTTAACCTCCACAGAACACA 1260
 |||||
 DB 1224 CTCGGGCTTTCAGTTAAACCAACCAATGCTTAAATCTCTTAACCTCCACAGAACACA 1283
 |||||
 QY 1261 GGACAGCCTCTTGTAGACGATGGGATTCGTTCAAGACTCTAGTTAATAGCTTTCAGANT 1320

```

Db 1284 GGACAGCCTCTGTAGACGATGGGATGCTTCAAGACTCTAGTAATAGCTTCAAGAAT 1343
Qy 1321 CACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGC 1380
Db 1344 CACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGC 1403
Qy 1381 ANATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATTGAACAGCTGTTGTCAGTAA 1440
Db 1404 ANATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATTGAACAGCTGTTGTCAGTAA 1463
Qy 1441 TGATTTGCARAACTATGATATTCGACTTTAAATAATCAAGCTTAATTCATATAAAC 1500
Db 1464 TGATTTGCARAACTATGATATTCGACTTTAAATAATCAAGCTTAATTCATATAAAC 1523
Qy 1501 TCGATGTAGAGATGGTGGTTCATGATACTACTTTTACAT 1540
Db 1524 TCGATGTAGAGATGGTGGTTCATGATACTACTTTTACAT 1563

```

```

RESULT 2
AY084227
LOCUS AY084227 1541 bp mRNA linear PLN 21-JUN-2002
DEFINITION Arabidopsis thaliana clone 100372 mRNA, complete sequence.
ACCESSION AY084227
VERSION AY084227.1 GI:21402937
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1541)
Haas,B.J., Volfovsky,N., Town,C.D., Troupkhan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1541)
Brover,V., Troupkhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1541)
Brover,V., Troupkhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
FEATURES
Location/Qualifiers
1..1541
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="100372"
59..1459
/codon_start=1
/product="vacuolar processing enzyme/asparaginyl
source
CDS

```

```

endopeptidase, putative"
/protein_id="AAMG0827.1"
/db_xref="GI:21536495"
/translation="MSSPLGHFOILVFLHALLIFSAESRKTOLLNDVSSDKSAKG
TRAVLVAGSNENYNYRQADICHAYOILRGGGLKDENLIVFVDDIADFSSNPPGV
IINKPDGEDYKGVPKDYTKAVNVQNFYNVLLNGESVGTGNGKVKYKSGNDNFIY
YADHAGPLGLIAMPTGDEVMARDFNEVLEKHKRIYKMKVIYVACSGSGMFGILKK
NLNIAYVATANSKSSSWGVCPEYPPPEISIGTCLGDTFISLWEDSLDLMGSKETL
EQOYHVYKRVGSDVPETSHVCEKMLKDYLSYIGRNPNDNFTFESFSPIS
NSGLVNPDRDPLPLYLORKIOKAPMGSLSEKAOKKLLDEKNHRKOIQDSTDIURLSV
KOTVNLNLLSTRITGLOPLVDWDFKTLVNSFKHCGATVHYGLKYTGALANICMG
VDVQKTSATEOAGSM"
BASE COUNT 499 a 287 c 321 g 434 t
ORIGIN
Query Match 97.4%; Score 1518.8; DB 8; Length 1541;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1520; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CTCACAAGATCAGATCAAGATAGAAATTTTCAACAAGATGTCTAGTCCTCTGGTGCAC 60
Db 20 CTCACAAGATCAGATCAAGATAGAAATTTTCAACAAGATGTCTAGTCCTCTGGTGCAC 79
Qy 61 TTTTCAGATTTCTTTTCTTCATGCTTTGCTTATCTTCTCAGCTGAGTCCGCGAAAC 120
Db 80 TTTTCAGATTTCTTTTCTTCATGCTTTGCTTATCTTCTCAGCTGAGTCCGCGAAAC 139
Qy 121 CAATTGCTGAACGATATGATGTTGAATCTAGCGACAGAGTGCAGAAAGGCACAGATGG 180
Db 140 CAATTGCTGAACGATATGATGTTGAATCTAGCGACAGAGTGCAGAAAGGCACAGATGG 199
Qy 181 GCTGTTTGTAGTTCGGATCAAAATGAATATATTAACACAGGCATCAGGCTGACATATGC 240
Db 200 GCTGTTTGTAGTTCGGATCAAAATGAATATATTAACACAGGCATCAGGCTGACATATGC 259
Qy 241 CACGCGTATCAGATACCTCCGAAAGGCGTTTAAAGATGAAGACATCATTTGTTGTTATG 300
Db 260 CACGCGTATCAGATACCTCCGAAAGGCGTTTAAAGATGAAGACATCATTTGTTGTTATG 319
Qy 301 TATGATGATATCGCGTTTCTTCGAGAAATCCTAGGCGCTGGAGTTATCATTAATAAACCA 360
Db 320 TATGATGATATCGCGTTTCTTCGAGAAATCCTAGGCGCTGGAGTTATCATTAATAAACCA 379
Qy 361 GATGAGAAAGATGTTTATAAAGAGATTCCTTAAGGACTACACTAAAGAGCTGTTAATGTT 420
Db 380 GATGAGAAAGATGTTTATAAAGAGATTCCTTAAGGACTACACTAAAGAGCTGTTAATGTT 439
Qy 421 CAAACCTTCTACATGTTTACTTGGAAATCAAACTGGCGTCCACAGGAGGAATGGCAAA 480
Db 440 CAAACCTTCTACATGTTTACTTGGAAATCAAACTGGCGTCCACAGGAGGAATGGCAAA 499
Qy 481 GTTGTGAAAAGTGGTCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAGCTCT 540
Db 500 GTTGTGAAAAGTGGTCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAGCTCT 559
Qy 541 GGCTTAATAGGATGCCCGACCTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGCTTTG 600
Db 560 GGCTTAATAGGATGCCCGACCTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGCTTTG 619
Qy 601 GAGAGATGCATAGAGAAAAAATACAAAGATGGTGATCTATGTTGAAGCATGTGAA 660
Db 620 GAGAGATGCATAGAGAAAAAATATACAAAGATGGTGATCTATGTTGAAGCATGTGAA 679
Qy 661 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACGAGTGAAGTCT 720
Db 680 TCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATACGAGTGAAGTCT 739
Qy 721 GCTAATTTCTAAAGAGACAGCTGGGAGATTTTACTGTCTCTAGTATATCTCTCTCTCT 780
Db 740 GCTAATTTCTAAAGAGACAGCTGGGAGATTTTACTGTCTCTAGTATATCTCTCTCTCT 799
Qy 781 TCTGAGATTTGGAACCTTGTCTCGGCGATACATTTAGCATCTCTTGCTTGGCTTGAAGAC 840

```

```

Db      800 TCTGAGATTGAACTTCTCTCGCGGATACATATTAGCATCTCTTGGCTTGAGCAGAGTGAC 859
QY      841 CTTTCATGATGACAAAGAGACTTTGGAGCAACAATACACAGCTTGTAAAGAGAGAGATGA 900
Db      860 CTTTCATGATGACAAAGAGACTTTGGAGCAACAATACACAGCTTGTAAAGAGAGAGATGA 919
QY      901 GGATCTGATGATGACAGAGACTTCTCATGTATGCGGTTTGGACAGAGAGATGCTTAA 960
Db      920 GGATCTGATGATGACAGAGACTTCTCATGTATGCGGTTTGGACAGAGAGATGCTTAA 979
QY      961 GATTATCTTCTCTTACATTTGAAGAAATCTGAAACGATTAACATTCACATTTTCACGGAA 1020
Db      980 GATTATCTTCTCTTACATTTGAAGAAATCTGAAACGATTAACATTCACATTTTCACGGAA 1039
QY      1021 TCCCTTTTCCACCAATCTCTAATCTTGGCTTGGTCAATCCGCGGATATTCCTCTGCTA 1080
Db      1040 TCCCTTTTCCACCAATCTCTAATCTTGGCTTGGTCAATCCGCGGATATTCCTCTGCTA 1099
QY      1081 TACCTCCAGAGAGATTTCAAAAGCTTCAATGGATCACTTGAAGCAAGAAAGCTCAG 1140
Db      1100 TACCTCCAGAGAGATTTCAAAAGCTTCAATGGATCACTTGAAGCAAGAAAGCTCAG 1159
QY      1141 AAGAAATTTGTTGACGAAAGATCATAGAAACAATCGATCAGAGCATTTACAGCAT 1200
Db      1160 AAGAAATTTGTTGACGAAAGATCATAGAAACAATCGATCAGAGCATTTACAGCAT 1219
QY      1201 CTGCGGCTTTCAGTTAAACAAACCAATGCTTTAAATCTCTTAACTTCCACAAAGAAACA 1260
Db      1220 CTGCGGCTTTCAGTTAAACAAACCAATGCTTTAAATCTCTTAACTTCCACAAAGAAACA 1279
QY      1261 GGACAGCCTCTTGTAGACGATTGGGATTCCTCAAGACTCTAGTTAATAGCTTCAAGAT 1320
Db      1280 GGACAGCCTCTTGTAGACGATTGGGATTCCTCAAGACTCTAGTTAATAGCTTCAAGAT 1339
QY      1321 CACTCGGCTCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTCCCAATATCTGTC 1380
Db      1340 CACTCGGCTCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTCCCAATATCTGTC 1399
QY      1381 AATATGGAGTGGATGTGAAGCAAACTGTTTCAGCAATGAACAAGCTTGTTCGATGTA 1440
Db      1400 AATATGGAGTGGATGTGAAGCAAACTGTTTCAGCAATGAACAAGCTTGTTCGATGTA 1459
QY      1441 TGATTTCCAAACAAATCTGATATTCGACTTTAAATAATATCAAGTTAATTTCAATAAAC 1500
Db      1460 TGATTTCCAAACAAATCTGATATTCGACTTTAAATAATATCAAGTTAATTTCAATAAAC 1519
QY      1501 TCGATGTAGAGATGGTGGTTC 1522
Db      1520 TCGATGTAGAGATGGTGGTTC 1541

RESULT 3
AP000383
LOCUS      82253 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MAL21.
ACCESSION AP000383 BA000014
VERSION    AP000383.1 GI:5672520
KEYWORDS
SOURCE
  Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
  clone:MAL21.
ORGANISM
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
  1 (sites)
REFERENCE
  Kaneko,T., Kato,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
  Structural analysis of Arabidopsis thaliana chromosome 3. II.
  Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
  and BAC clones
  DNA Res. 7 (3), 217-221 (2000)
  20363099
REFERENCE
  Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  (bases 1 to 83253)
  20363099
REFERENCE
  Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  (bases 1 to 83253)
  20363099

```

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yane,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MAL21
Genes with similarity to proteins in the databases are described in
'product', or 'note', qualifiers. Genes that have no significant
product similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://combio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremblin.zozi.lastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MZEL9 and the 3' clone is MQCL2.
Location/Qualifiers
1..83253
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MAL21"
/clone_lib="Mitsui P1"
complement(1..376)
/product="nucleoid chloroplast DNA-binding protein-like"
/note="CDS is reported in Acc# AP002050
gene_id:MZEL9.7"
/number=1
/evidence=not_experimental
complement(join(4273..4315,4392..4510,4727..4902,
5273..5377,5836..5914,6019..6096,6222..6390,6555..6620,
7231..7266,7418..7486,7659..7724,7870..8163))
/note="gene_id:MAL21.2"
/codon_start=1
/evidence=not_experimental
/product="protein arginine N-methyltransferase-like
protein"
/protein_id="BAB01859.1"
/db_xref="GI:9293936"
/translation="MQSGDFNSGFGHDHRELELEDKOGPSLSSFGRAKRSKSHAGAR
DPRGLANLVRSVDQLGEHKSLETSPPPTDFDVAFYHSAHVGIHEEMIKDRAR
TETRYREAIHQHSLIEGKVVVDVCGTGILSIFCAQAGAKRVYAVDASJAVQAKREV
KANGLSKVTVLHGRVEDVEIDEDVIISENMGYMLLYESMLGSVITARDRLKPGV
LILPESHATYMAPISHPDYSHSIDFNRNVYIDMSAMQAKQCAFEPSPVESISGE
NVLTWPVFPFLFSGPSPAPKATSETSIASGSSSISPSGEVKNQKRTNPSDALVLS
TSPSPPTTHMQQTIVFYDPIDVEDQDQIEGVTLSQSKENKRFNHNHLEYSACKRSF
VKESVMR"
10472..11680
/note="emb|CAB78723.1
gene_id:MAL21.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01860.1"
/db_xref="GI:9293937"
/translation="MTMMSDLSDLLLEETLSBVPRTSLCAVRSCKRWNTLFKDRILC
KARETDQRFIMKKYKLSMSRFDLNGTLNEGGTFEDVPSIKELGHFNQVKISKVF
QCGLLCTVKEDNIRLKYLNPNFLGQIRIESGNTNRLFDYRIGYDNNRKHILRF
LEYINFMKHFLEYIEITDSSNRVRLDIAPRWELTESYQFGASLKGNFTYFAKRIE
YEEGDFPEPADNLLCTFTTSEFGCLPLFPQHYLDVGLUSSLDKLAALFOCGD
TDLSEVEIVTDTLTETNYSWNPFLKVDMEPHYGRSFMFDYGGSFIDEEKKLAVF
OFDESGMTRYEDTATYIIGENGYYKVRLEGEAFANOGGYCFPSVCSFSSVYPSLVQINQI
VGPKKEREE"

FEATURES
source

exon

CDS

CDS


```

CDS
join(15535..15806,16005..16155,16253..16435,16674..16748,
16945..17100,17229..17327,17425..17552,17741..17819,
17917..18282)
/note="gene_id: MAL21.4"
/codon_start=1
/evidence=not_experimental
/product="chloroplast outer envelope hexokinase"
/protein_id="BAB01861.1"
/db_xref="GI:9293958"
/translation="MGVLYMLTAAAVACSVATVVRMRKRRKWRVVGLLKDL
EACETPLGRUOWDAIVAMQAGLVSEGGSKLMLLTFVDDLPNGSETTYVALHL
GGSYFRIKHLGORSLEQVDVERHSIPTSLMNSTSEVLFELASLSORFIEKEGN
DFSLSQPLKRELATTFPPYKQTSISSGYLKWTKGRAISEMAGEDIAECLQALNKR
GLDRLAALVNDTVALSGFHDPTTAAVFGTSGNACYLERTDAILKQCNPTTS
GSMYNNMGWNGFSRPLRSTYSIDLEDAESMNSDMGFEMIGMYLGDIVRVYILRM
SQESDIPTGSIPTSLFTVLRNVSAMHEDDTSELOEARILKDLGVSEYPMKVRKL
VVKICDVTRRAALAAAGIAGILKVGCRDGGGRRSDKOIMRRTVVAVEGGLYNY
RMFREYDEALRDLGEDVAQHVVVKAMEDGSSIGSALLASSOSVOTIPSV"
complement(join(18762..18845,18933..19038,19118..19221,
19306..19434,19520..19624,19780..19914,19994..20080,
20208..20304,20350..20767,20863..20928,21116..21211,
21364..21414,21501..21560,21669..21755,21868..21957,
22170..22234,22412..22484))
/note="gene_id: MAL21.5"
/codon_start=1
/evidence=not_experimental
/product="chaperonin, t-complex protein alpha subunit"
/protein_id="BAB01862.1"
/db_xref="GI:9293959"
/translation="MSISAONPDISGDRSQGVYRTONVMCAQVSNIVKTSLGPVGL
DKMLVDIGDVTINDGATILRMLEVEHPAKVYLAELODREVGGTTSVIVAAE
LLKRLANDLRNKIHTPISIGYRLAMRESCKYIEELKVLTKVEKLGVPGLINCAKTSMS
SKLTSQSDSFANLVPAKISVKNQGEIKYPIKGINILKAHQSGARDSYLLNGYE
INTGRAAGMPLRVSPAKIACLDENLQKTMQLGVVVVNDPRLKIRQREADMTKE
RIEKLKAGANVITTKGIDDMALKYFEAGAIARRVRKEDMRHVAKATCATVTFE
ADMGEETOPAHLSGDAEVEVEERADDDVILIKGKTSSTSAVSLILRGANDYMLDEME
RALHDALCIKVTILESNTVAGGAVESALSYLEHLATILGSREQLAIAEFADALLI
IPKVLVAANKADTELVAKLRAHYHTAQTKADKKHYSRKKRCSMGLDVLNGTIRNNLE
AGVIEPAMSKVITIOFATEAAITILRIDDDMKILVKDESQEE"
complement(join(23145..23191,23289..23348,23451..23514,
23623..23850,23934..23961,24383..24501))
/note="gene_id: MAL21.6"
/codon_start=1
/evidence=not_experimental
/product="ubiquitin conjugating protein-like"
/protein_id="BAB01863.1"
/db_xref="GI:9293960"
/translation="MATVNGYTGNTPAATPAATGSKOSAPPTKTVDSHSLVKLPLOSE
LMGLMGADPGISAFPEEDNIFCWKGTITGSKDVPFGTEYRLSLTFSNDYPPFKSPRV
KFETCCHPNVDLGNICLDILQDKWSAYDVRTILLISLLEGPNISPLNQAQO
LMSNQEEYRKMVEKLYPLNA"
complement(join(25168..25358,25476..25644,25764..25981,
26173..26234,26335..26425,26620..26689,26796..26843))
/note="gene_id: MAL21.7"
/unknown_protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01864.1"
/db_xref="GI:9293961"
/translation="MEALYAKLYDKYTKLOKKYSEYDEINKEOEFKPLTFVSAEEL
MEHLRGENOSLEWVKLRNELIISGRDDKFLKCOLLMEEELKNKSISEVVVKL
ELVEEHPRNVDSGKKRKRKPESARVTRSMKRSRUSEDLVDMDVSPDSKHH
KARLEPLVSPQCCRTYTDGSSASCTFOALGKHLMLKSLTNKKKACIVASHPT
TGLSFSLFTFINNPNGESELLYPKPSLGTQFVAPENWREVIKFTSMCPIFFERVSRR
VIKLCN"
27734..27973
/note="contains similarity to cytochrome P450"
gene_id: MAL21.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01865.1"
/db_xref="GI:9293962"
/translation="MMLQCFDWKTKGDKVDNEEAIRGLNLTMAHPLKCIPIVARNLNL

CDS
ACSNLKPONFEFAKKLLVRGSSAPLCMPFPCPSLPV"
join(29031..29951,30031..30681)
/note="gene_id: MAL21.9"
/codon_start=1
/evidence=not_experimental
/product="cytochrome P450-like protein"
/protein_id="BAB01866.1"
/db_xref="GI:9293963"
/translation="MAAMNIVDFQNCFFVLLSLFSLCYSLFFRKPKPELRHLYDLP
SPPTPIGHLHLLSLVHRSLOTSTKYGSLVLRVFSFVALVSSASTAYETFE
HDVNSISRGPPPTDLSLFAFSFTSPAPGYDMFKMKLLVTLNLLGPOLERSGFR
DELDTFENLDDKAMKESVDICVEALKSNNSCKIMGSCSEENGEAEVRALAT
QDGLTKRILLUANRAGFKLVSLFRKMMDSVRSDLEULRILYVEHDKLDHMQ
GTDLDALLAACROKNAEYKTSRNHISFFADLLFASDTDTQVOTWVAEIIINNPN
LERUGEDSVKARLTQETDNLPLYQAVVKEGRLHPGLFARFSQEGGRIGG
FYRPEKTTMINAVMRDSDSDWEDPDEPKPERFLASSRSEQEKERRQAIKYTAFGS
FRRSCPENLAYIFLGTAGVMOGFEWRIKEEVNMEANVLSLTWAYPLKVTVP
RTLVLPTQNLNRS"
join(32158..32475,32844..33086,33190..33813)
/note="contains similarity to cytochrome P450"
gene_id: MAL21.10"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01867.1"
/db_xref="GI:9293964"
/translation="MAAMINVEFQNCIFIMLLCLFSLCYSLFFRKPPSSRQNSYDCDL
PPSPSPMPIIGHLHLISLVLHKSQFKISSYKQGLLHLRINFNPVILVSSASVADIF
RVHDKFFLAATLRLLEKIGLSLFOKEIMGVSRGFDLELLRILREHEBKPDHHDMD
MDALLAAYNDKAEYKTRNQIKAFIVDMFIAGTDISALTTQGTMAEILNPNVLVRI
REKIDSVVGRSLRLOETDLPKLPYSQAVVKEGRLHPPTLUMVREFQGGCKVGFYIP
ASTTLVNGYAVMRDNPWEEPEFKPERFLASSRSEEEIREQALKYIAFGSGRRG
CPGNTIAYIFVGTAGMVQCFDKIRGDKVDMKEAIGGLNLTLAHPLKCTPVARSPP"
join(34864..34912,34957..35241,35310..35443,35870..36133)

CDS
Query Match 53.3%; Score 832; DB 8; Length 83253;
Best Local Similarity 70.6%; Pred. No. 1.4e-178;
Matches 1545; Conservative 0; Mismatches 0; Indels 643; Gaps 7;

Qy 1 CTCACAAGATCAGATTCACAGATGAGCTTTTCAACAACATGTCTAGTCCTCTTGGTCCAC 60
Db 72661 CTCACAAGATCAGATTCACAGATGAGCTTTTCAACAACATGTCTAGTCCTCTTGGTCCAC 72720

Qy 61 TTTCCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTTCTCAGCTGAGTCCCGCAAAACC 120
Db 72721 TTTCCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTTCTCAGCTGAGTCCCGCAAAACC 72780

Qy 121 CAATTCGTGACGATATGATGTTGAATCTAGCCACAGAGTGCNAAAGGCACACGATGG 180
Db 72781 CAATTCGTGACGATATGATGTTGAATCTAGCCACAGAGTGCNAAAGGCACACGATGG 72840

Qy 181 GCTGCTTTAGTTGCTGGATCAAAATGAATATATTAACACTACAGCAT----- 225
Db 72841 GCTGCTTTAGTTGCTGGATCAAAATGAATATATTAACACTACAGCATGTTGTTAAATTA 72900

Qy 226 ----- 225
Db 72901 TGTTTGAACGTTTAAACATAACAAAAAAGGTCCCAACGAGATTTGTATGAACATA 72960

Qy 226 -----CAGCTGACATATGCCACGCGTATCAGATACTCGGAAAA 264
Db 72961 ATCGACGACGCTTTTATTTTCACAGGCTGACATATGCCACGCGTATCAGATACTCGGAAAA 73020

Qy 265 GGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATATCGCGTTTCTCTCG 324
Db 73021 GGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATATCGCGTTTCTCTCG 73080

Qy 325 GAGAACTCTAGGCGCTGGAGTTTATCATTAAATAACACAGATGAGAGAGTGTATATAAGGA 384
Db 73081 GAGAACTCTAGGCGCTGGAGTTTATCATTAAATAACACAGATGAGAGAGTGTATATAAGGA 73140

Qy 385 GTTCCTA----- 391
Db 73141 GTTCCTA----- 391

```

Db 73141 GTTCCTAAGGTTCTTATTTCTACTCTTTTGTGGTTATTTCTACGTTGAATTCAAATAC 73200
QY 392 -----AGGACTACACTAAGAAGCTGTGA 415
Db 73201 ATATATATATCAAGTTTTTGTGTATTTGGTTAGGACTACACTAAGAAGCTGTGA 73260
QY 416 ATGTTCAAAAGCTTCTACAATGTGTTACTTGGAAATGAAAGTGGCGTCACAGAGGAAATG 475
Db 73261 ATGTTCAAAAGCTTCTACAATGTGTTACTTGGAAATGAAAGTGGCGTCACAGAGGAAATG 73320
QY 476 GCAAAGTTGTGAAAGTGGTCCCTAAATGATAATATCTTCACTATATATGCTGACCATGGAG 535
Db 73321 GCAAAGTTGTGAAAGTGGTCCCTAAATGATAATATCTTCACTATATATGCTGACCATGGAG 73380
QY 536 CTCCTGGCTTAAT----- 548
Db 73381 CTCCTGGCTTAATAGGTTTTCTTAATTTTATGAAATTTATACGTACCATAATCCATATC 73440
QY 549 -----AGCGATGCCACATG 562
Db 73441 TATAATAAAGATTTTCTCTTGATACACGAACCGGATTTTCTCAGCGATGCCACATG 73500
QY 563 GTGATGAAGTTATGCGCAAAAGATTTCAATCAAGCTTGGAGAAGATGCATAGAGAAAA 622
Db 73501 GTGATGAAGTTATGCGCAAAAGATTTCAATCAAGCTTGGAGAAGATGCATAGAGAAAA 73560
QY 623 AATACACAGAT----- 635
Db 73561 AATACACAGATGGTATATACTCAACCATTCGTTACCTAGCTTTATACATATGTGTTTC 73620
QY 636 -----GGTGATCTATGTTGAAGCATG 656
Db 73621 TGTTTTTGAATCTATGGTGTGTTTTTTGGATGTTTGGTGATCTATGTTGAAGCATG 73680
QY 657 TGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATATGCGCAGTGAC 716
Db 73681 TGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTCAACATATATGCGCAGTGAC 73740
QY 717 TGCTGCTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCTCTCTCC 776
Db 73741 TGCTGCTAAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCTCTCTCC 73800
QY 777 TCCTTCTGAGATTTGGAATCTGCTCGCGGATACATTTAGCATCTCTTGGCTTGA----- 830
Db 73801 TCCTTCTGAGATTTGGAATCTGCTCGCGGATACATTTAGCATCTCTTGGCTGAGGACAG 73860
QY 831 ----- 830
Db 73861 GTACTGCAACAAAAAGATTCATCTTATGCACTATTCGAATGATTTGTTCTT 73920
QY 831 -----GGACAGTGAACCTTCATG 847
Db 73921 GAAGAATATTTGTTCAATTTCTTATGTTTTGTGTGTTTGGGACAGTGAACCTTCATG 73980
QY 848 ACATGAGCAAGAGACTTTTGAGCAACAATPACCAGCTTGTAAAGAGAGAGTAGGATCTG 907
Db 73981 ACATGAGCAAGAGACTTTTGAGCAACAATPACCAGCTTGTAAAGAGAGAGTAGGATCTG 74040
QY 908 ATGTACCAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAGATGCTTAAAGATTATC 967
Db 74041 ATGTACCAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAGATGCTTAAAGATTATC 74100
QY 968 TTTCTCTTACATTTGAAGAAATCCTGAAACGATAACTTCACTTTCACCGGAATCCTTTT 1027
Db 74101 TTTCTCTTACATTTGAAGAAATCCTGAAACGATAACTTCACTTTCACCGGAATCCTTTT 74160
QY 1028 CCTCACCATCTCTTAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCC 1087
Db 74161 CCTCACCATCTCTTAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCC 74220
QY 1088 AGAGAA----- 1093
Db 74221 AGAGAAAGGTGAGCTTTTTCGGGTTTTTGTGATCATTTTAAACGAAGAGATTTTCAGCAT 74280

QY 1094 -----AGATTTCAAAAAGCTCCAAATGGATCAGTCTTGAAGCAA 1130
Db 74281 GTTTTAATATTTATTCATCTCTTAGATTTCAAAAAGCTCCAAATGGATCAGTCTTGAAGCAA 74340
QY 1131 AGAAGCTCAGAAAGAAATTTGTTGACGAAAAGAAATCATAGGAAAACAAATCGATCAGAGCAT 1190
Db 74341 AGAAGCTCAGAAAGAAATTTGTTGACGAAAAGAAATCATAGGAAAACAAATCGATCAGAGCAT 74400
QY 1191 TACAGACATTTCTGGCGCTTTTCAGTTTAAACAAACCAATGCTCTTAAATCTCTTAACTTCCAC 1250
Db 74401 TACAGACATTTCTGGCGCTTTTCAGTTTAAACAAACCAATGCTCTTAAATCTCTTAACTTCCAC 74460
QY 1251 AAGAACACAGCAGCAGCTCTTGTAGACGATTCGGATTTGCTTCAAGACTCTA----- 1302
Db 74461 AAGAACACAGCAGCAGCTCTTGTAGACGATTCGGATTTGCTTCAAGACTCTAGTAAACAA 74520
QY 1303 ----- 1302
Db 74521 CCACATCTCAACCTTTGTTACTTGTCTTACGCAACCAACCATTCGATTTACTATAACCC 74580
QY 1303 -----GTTAATAGCTTCAAGAATCACTGGGTGCAACGGT 1337
Db 74581 AGTGATATCGAATGAAATCGCAGGTTAATAGCTTCAAGAATCACTGGGTGCAACGGT 74640
QY 1338 GCATTACGGATTTGAAGTATACAGAGCGCTTGGCAATATCTGCAATATGAGGAGTGGATGT 1397
Db 74641 GCATTACGGATTTGAAGTATACAGAGCGCTTGGCAATATCTGCAATATGAGGAGTGGATGT 74700
QY 1398 GAAGCAAACTGTTTTCAGCCCATTTGAACAAAGCTTGTTCGATTAATGATTTTGCAAAACAAATG 1457
Db 74701 GAAGCAAACTGTTTTCAGCCCATTTGAACAAAGCTTGTTCGATTAATGATTTTGCAAAACAAATG 74760
QY 1458 TGATATTCGACTTTTAAAAATATCAAGTTAAATTTCAATAAACTCGATGTAGAGATGGTT 1517
Db 74761 TGATATTCGACTTTTAAAAATATCAAGTTAAATTTCAATAAACTCGATGTAGAGATGGTT 74820
QY 1518 GTTTCATGATACTACTTTTACATGAAA 1545
Db 74821 GTTTCATGATACTACTTTTACATGAAA 74848

RESULT 4
AY070375 556 bp mRNA linear PLN 22-APR-2002
LOCUS Arabidopsis thaliana At3g20210 mRNA sequence.
DEFINITION Arabidopsis thaliana
ACCESSION AY070375
VERSION AY070375.1 GI:17979293
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 656)
Yamada, K., Banh, J., Chang, C.H., Chang, E., Dale, J.M.,
Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.,
Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 656)
Yamada, K., Banh, J., Chang, C.H., Chang, E., Dale, J.M.,
Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.,
Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,

Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
 Direct Submission
 Submitted (18-DEC-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.
 The Salk, Stanford, PGECC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada, K., Banth, J.,
 Chang, C.H., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M.,
 Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,
 Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T.,
 Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M.,
 Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W.,
 Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
 contributed equally to this work as PIs.

FEATURES
 Source
 1..656
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL09-59-E07 (R19019)"
 /note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert.
 ecotype: Columbia"
 1..656
 /gene="At3g20210"
 63..647
 /note="At3g20210"
 /note="This is a partial cDNA containing no stop codon.
 However, it does have a polyA tail which is not present in
 the genomic sequence. This cDNA is a partial
 correspondence to gene At3g20210."
 /evidence="experimental"
 BASE COUNT 221 a 113 c 135 g 187 t
 ORIGIN

Query Match 40.2%; Score 627.2; DB 8; Length 656;
 Best Local Similarity 99.5%; Pred. No. 6.e-132;
 Matches 629; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CTCACAGAATCAGATTCAGATAGAGTGTTCACAAATGCTAGTCTCTTGGTTCAC 60
 DB 24 CTCACAGAATCAGATTCAGATAGAGTGTTCACAAATGCTAGTCTCTTGGTTCAC 83
 QY 61 TTTCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTCTCAGCTGAGTCCCGCAAAACC 120
 DB 84 TTTCAGATCTTGTGTTTCTTCATGCTTGTGCTTATCTCTCAGCTGAGTCCCGCAAAACC 143
 QY 121 CAATGCTGACGATATGATGTTGAATCTAGGACAAAGTGCACAAAGCAGCATGG 180
 DB 144 CAATGCTGACGATATGATGTTGAATCTAGGACAAAGTGCACAAAGCAGCATGG 203
 QY 181 GCTGTTTTAGTCTGATCAATGAATATATTAACCTACAGGATCAGGCTGACATATGC 240
 DB 204 GCTGTTTTAGTCTGATCAATGAATATATTAACCTACAGGATCAGGCTGACATATGC 263
 QY 241 CACGCGTATCAGATCTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTTTATG 300
 DB 264 CACGCGTATCAGATCTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTTTATG 323
 QY 301 TATGATGATATCGGTTTCTCGGAGATCTTAGGCTCGAGTTATCATTAATAACCA 360
 DB 324 TATGATGATATCGGTTTCTCGGAGATCTTAGGCTCGAGTTATCATTAATAACCA 383
 QY 361 GATGGAGAAGATGTTTATAAGAGTTCCTAAGGACTACACTAAAGAGCTGTTATGTT 420

Db 384 GATGGAGAAGATGTTTATAAAGAGTTCCTTAAGGACTACTAAAGAAGCTGTTAATGTT 443
 QY 421 CAAAACCTTCTACAAATGTTTACTTGGAAATGAAAGTGGCTACAGGAGGAAATGGCAAA 480
 Db 444 CAAAACCTTCTACAAATGTTTACTTGGAAATGAAAGTGGCTACAGGAGGAAATGGCAAA 503
 QY 481 GTTGTGAAAGTGGTCTTAATGATATATCTTATCTATCTATCTATCTATCTATCTATCT 540
 Db 504 GTTGTGAAAGTGGTCTTAATGATATATCTTATCTATCTATCTATCTATCTATCTATCT 563
 QY 541 GGCTTAAATAGCGATGCCACTGCTGATGAAGTATGCAAAAGATTTCAATGAAGTCTTG 600
 Db 564 GGCTTAAATAGCGATGCCACTGCTGATGAAGTATGCAAAAGATTTCAATGAAGTCTTG 623
 QY 601 GAGAAGATGCATAAGAGAAAAAATAACAACA 632
 Db 624 GAGAAGATGCATAAGAGAAAAAATAACAACA 655
 RESULT 5
 AF238384
 LOCUS 1576 bp mRNA linear PLN 02-MAR-2001
 DEFINITION Vigna radiata asparaginyl endopeptidase (PEL) mRNA, complete cds.
 ACCESSION AF238384
 VERSION AF238384.1 GI:13183094
 SOURCE Vigna radiata.
 ORGANISM Vigna radiata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Vigna.
 REFERENCE 1. (bases 1 to 1576)
 AUTHORS Yang, W.H. and Chen, C.S.
 TITLE Vigna radiata mRNA for asparaginyl endopeptidase, complete cds
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1576)
 AUTHORS Yang, W.H. and Chen, C.S.
 TITLE Direct Submission
 JOURNAL Submitted (24-FEB-2000) Institute of Botany, Academia, Sinica,
 Nankang, Taipei, Taiwan 11529, Republic of China
 FEATURES
 source
 1..1576
 /organism="Vigna radiata"
 /cultivar="Tainan 5"
 /db_xref="taxon:157791"
 /tissue_type="cotyledon"
 /dev_stage="germinating seeds"
 1..1576
 /gene="PEL"
 19..1470
 /gene="PEL"
 /codon_start=1
 /product="asparaginyl endopeptidase"
 /protein_id="AAK15049.1"
 /db_xref="GI:13183095"
 gene
 CDS
 BASE COUNT 419 a 327 c 377 g 453 t
 ORIGIN
 Query Match 30.0%; Score 467.6; DB 8; Length 1576;
 Best Local Similarity 60.8%; Pred. No. 1.e-95;
 Matches 780; Conservative 0; Mismatches 499; Indels 3; Gaps 1;
 QY 150 TAGCGACAAGAGTGCAAAAGGCACACGATGGGCTGTTTTAGTGTGCGATCAATGAATA 209

	Qy	Qy	Db	Qy	Db
	1351	AAGTATACAGGAGCGCTTCCCAATATCTGCCAATATCGGGAGTGGATGATGGAAGCAAACTGTT	1410		
	1370	AAACACATGCGTTCGATTTGCAACACATGATGCAATGCCGGATCAACAATAATCAATGGCC	1429		
	1411	TCAGCCATTGAACAAGCTTGTTCGA	1435		
	1430	GAGGCTTCAGCACAAAGCTTGCOCGA	1454		

Sequence

RESULT	7
D89972	
LOCUS	D89972 1659 bp mRNA linear PLN 06-APR-1999
DEFINITION	Vigna mungo mRNA for asparaginyl endopeptidase (VMP2-LA), complete cds.

ACCESSION	D89972
VERSION	D89972.1
KEYWORDS	GI:4589397
SOURCE	asparaginyl endopeptidase (VmPE-1A).
ORGANISM	Vigna mungo seedling cotyledon cDNA to mRNA, clone:pPEB.
	Vigna mungo
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.

REFERENCE	AUTHORS	TITLE
1 (BASES 1 TO 1039)	Okamoto, T. and Minamikawa, T.	Molecular cloning and characterization of Vigna mungo processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-PP)

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Plant Mol. Biol. 39 (1), 63-73 (1999)
99178794
2 (bases 1 to 1659)
Okamoto T.
Okamoto T.
Direct Submission
Submitted (19-DEC-1995) Takashi Okamoto, Tokyo Metropolitan
University, Department of Biology; Minami-osawa 1-1, Hachioji,
Tokyo 102-03, Japan (E-mail:okamoto-takahashi@metr.u.ac.jp,
Tel: +81-426-77-2562, Fax: +81-426-77-2559)

FEATURES	source
5'UTR	Location/Qualifiers 1. .1659 /organism="Vigna mungo" /db_xref="taxon:3915" /clone="pPEB" /tissue_type="cotyledon" /dev_stage="seedling" 1. .45 /evidence=experimental
CDS	46. .1494 /note="processing enzyme 1A" /codon_start=1 /evidence=experimental /product="asparaginyl endopeptidase (VmPE-1A)" /protein_id="BAA76745.1" /db_xref="GI:4589398"

Q7GRWAILFAGSNGWNYRHQADICHAYOILTRKGGLKEENIIVFM7DDTLSPDNDNPRP
/trans/...=MATTYSLSLFLFLATVALVAARHDHVGDGLFSLDSDSCNDNDI
G7LIIKRPDGDVVEGVKDYTGEDATAHNFYSALLGDKSALTCGSKSVSSGSGPDDRLL
IFYSDHGGPVLGTGPAGTYIASDLVEVLKKHASCTYKNLVEYLACAEAGSTFEGELL
PEDINIYATASNAEBSNGTFCGTEPPPEPTSTCLGDLVSANAEBSDRINLRTE
SLHQQYVVKORDADLVFHWKHYQGVGDVEFSKDALFLYLGTDPNANDTLFVDSNSLW
SS2TAYNORDADLVFHWKFRKAPGSKNEARKQVLEYMSHRMIIDDSVKLVGLKLL
FGFEKAPVLYNVRPAGSALVDWQACLKTMVTRTFETHCGSLSQYGHKHMSPFANICNV
GIKKKEOWAEASQACVTVTPASSNSLORGEFA

3'UTR	1495..1659	/evidence=experimental	
BASE COUNT	446 a	344 c	390 g 479 t
ORIGIN			
Query Match	29.9%	Score 466	DB 8: Length 1659

Matches 779; Conservative 0; Mismatches 500; Indels 3; Gaps 1;
Best Local Similarity 60.8%; pred. No. 2.4e-95;

150 TAGGACAGAGTGCACAAAGGACACAGATGGCTGTTTGTAGTCTGGATCAATGAATA 209
162 TAACGATGACAGTCCAGGAGCGCTGGCCATCTCTTCGCCGGTTGCAACGGCTA 221
210 TTATTAACATACAGGATAGGCTGACATATGCCAGCGGTATCAGATATCCGAAAGCGG 269
222 TTGGAACTACAGACACAGCGCGATATTTGTATGCTATCAATATAGGAAAGTGG 281
270 TTAAAAGATGAACATCATTTGTTTATGTATGATATATCCGCTTTTCTCGGAGAA 329
282 CTTGAAGAAAGAAATATCATTTGTTTATGTATGATGACATTCATTCAATTGGGACAA 341
330 TCTAGGCTGAGTATCATTAATAACACAGATGGAGAGATGTTTATAAGAGATTCC 389
342 TCCAAAGGCTGAGTATCATTAATAACACAGAGCGGGATGATCTTTATGAAGAGTTCC 401
390 TAAGACTACACTAAAGAGCTGTTTAAATGTTCAAAAGTCTTACAAATGTTTACTTGGAAA 449
402 AAAGATTACACTGGCGAAGATGCTACTGCTCAATACTTTTACTCTGTTTACTTGGAGA 461
450 TGAAGATGGCTGACAGAGGAAATGCAAGTTTCTGAAGTGGTCTCTAATGATATAT 509
462 TAAGTCGGCGCTTACCGGTGGCAGTGGGAAGTTGTGAGCAGTGGCGCCGATGATCGCAT 521
510 CTTTATCTATTATGCTGACAGTGGCTTCTGAGTCTTAATAGCGATGCCACTGGTGATGA 569
522 ATTATATCTACTCTGATCATATGTTGTCAGGGGTTCTCGGAGCGCTGCTGCTGCTTAA 581
570 AGTTATGCAAAAGATTTCAATGAAGTCTTGGGAAGATGATGAAGAAAGAAATACAA 629
582 CATATATGATCTGATTTGTTGAGTACTGAAGAAAGCAATGCTTCTGGAACGTATAA 641
630 CAAGATGGTATCTATGTTGAAGTATGTAATCAGGAAGTATGTTGAAGGATTTTAA 689
642 AAACCTAGTATTTATTTGAGGAGTGTGAAGCTGGAAGTATCTTCGAAGCTCTTCTTC 701
690 GAAATCTCAACATATACGAGTACTGCTGCTTAATCTAAAGAGACAGCTGGGAGT 749
702 TGAAGATATCAATATTTATGCAACAGCGCTTCAATGCGCAAGAAAGTAGTTGGGGAAC 761
750 TTACTGCTCTGAGTATATCT 809
762 ATATTGCTCTGGGAGTATCT 821
810 ATTTAGCATCTCTGGCTTGAAGAGAGTAGTAGTCTGATGATACAGAGACTTCTCATGT 929
882 CCAACATATTAAGTGGTTAGGATAGGACTCTATCTGGAGTGGTATGCGCTCTCAGT 941
930 ATGCGGTTTCGACAGAGAGATGCTTAAAGATATATCTTCTCTCTTACATTCGAGAAA 989
942 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1001
990 TCTGAAACGATATCTTCTTACGGAATCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1049
1002 TCTGCAATATATTTGATTTCTGTTGATGATGATGATGATGATGATGATGATGATGATG 1060
1050 CTTGCTCAATCCGCGGATATCT 1109
1061 --CAGTCAACAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1118
1110 AATGGATCTCTGAAAGCAAGAGCTCAGAGAAATTTGCTTGACCAAAAGATCATAG 1169
1119 GGAGGTTTCTCTAAGAAATTAAGCTCGGAACAAAGTTCTGGAAGTATATGCTCACAG 1178
1170 GAAACAAATCCATCAGAGCATTTACAGACTTCTCGGCTTTCAGTTAAACAAACCAATGT 1229
1179 AATGCATATACGACAGTGTGAACCTTTTGTGAAGAGCTTTTATTTGGATTTCAAGAGC 1238
1230 CTTAAATCTCTTAATCTCCACAAAGCAACAGGACAGCTCTCTCTAGACGATTTGGGATTC 1289

1239 TCCTGAAGTACTGAATGCTTTAGACCGCGGTGATCGGCACCTTGTGTAGTACTGGGCTG 1298
1290 CTTCAAGACTCTAGTAAATAGCTTCAAGAATCACTGCGGTCCAAACGGTGCATTACGAT 1349
1299 CTTGAAGACCATGGTAAGGACTTTTGAGACACATTTGTGATCTCTCTCAATGATGAT 1358
1350 GAAGTATACAGAGCGCTTGCATATCTGCAATATCTGCAATATGAGAGTGGTGAAGCAAACTGT 1409
1359 GAAACACATCACTGCTCTTTCGCAATATCTGCAATATGAGCAATAAAGAAAGCAAAATGCG 1418
1410 TTCAAGCCTTGAACAAAGCTTGT 1431
1419 TGAAGCCTCAGCAACAGCTTGT 1440

Sequence Variants

CSYSPRNS 1839 bp mRNA linear PLN 01-FEB-1996
LOCUS Cysine protease mRNA for putative cysteine proteinase.

DEFINITION 247793
ACCESSION 247793.1 GI:633184
VERSION 1
KEYWORDS Cysteine proteinase.
SOURCE Citrus sinensis.

ORGANISM

Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.

Alonso, J. and Granell, A.

1 (bases 1 to 1839)
A putative vacuolar processing protease is regulated by ethylene
and also during fruit ripening in Citrus fruit

Plant Physiol. 109: 5411-5447 (1995)

2 (bases 1 to 1839)

Alonso, J.

Direct Submission

Submitted (17-JAN-1995) Jose Miguel Alonso, Desarrolo Vegetal,
Inst. de Biologia Molecular y Celular de Plantas, CSIC-UPVA, Camino
de Vera 14, Valencia, Valencia, 46022, SPAIN

FEATURES

Source

1..1839
/organism="Citrus sinensis"
/db_xref="taxon:2711"
/clone="Tar4N1"
/tissue_type="Flavado"
/clone_lib="Orange Flavado E12 cDNA library"
/dev_stage="ethylene treated fruits"
19..1503
/codon_start=1
/product="cysteine proteinase (by similarity)"
/protein_id="CAA87720.1"
/db_xref="GI:633185"
/db_xref="SWISS-PROT:P49043"

CDS

1..1839
/translation="MRLASGLVLLVLAAGIADSRDIADGILKLPSEAYRFFHNG
GGAKVNDSDSVGRWAVLAGSGFMVNRHQADICAYOLLKRGKLDENIIVPMY
KDIAFNEENPRPGVLIINHPGDDVYKGVKDYTGEDYVEKFAVLGNKLTALTGSG
KVDSPGNDHIFIFYSDHGGPVLGMPTSRVYIADELIDVKKHAGSNYSKLVFYLE
ACESGIPGLLEGLNATYATASNAESSGYTCPEIPCPPEYSCICLDLYLAW
MEDSDIHLNRTETLHQOVELYKTRTASNSYSGVMQYGDIGISKNNLFTYLGTPAN
DNYTFVDSNLRPASKAVNQADLLHFWDKYPKAPGCTPRKAEQKQFPEANSHRMH
VDHSTKLIGKLPGLIEKPEILNTRPAGQPLVDDMGKLSKLVRTFESHCCALSOYCM
KHMRLANICNTGICKKMAEASAOACENIPSPWSSLDKGFSA"

BASE COUNT 499 a 368 c 436 g 536 t

ORIGIN

Query Match 29.7%; Score 463; DB 8; Length 1839;
Best Local Similarity 61.0%; Pred. No. 1, 1e-94;
Matches 770; Conservative 0; Mismatches 490; Indels 3; Gaps 1;

169 GCGACCATGGCTGTTTGTAGTCTGCGATCAATGATATTAATCACTACAGCATCAG 228
190 GGCACGATGGCGCTCTCTGCGGATCTATGCTTCTGGAATTTACAGGCATCAG 249

COMMENT	The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J., Hayashizaki,Y. and Shinozaki,K.	
	The Salk, Stanford, PGECC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K., Banh,J., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Yu,S., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.	
FEATURES	Yamada,K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC) contributed equally to this work as PIs.	
	Location/Qualifiers	
source	1..1516	
	/organism="Arabidopsis thaliana"	
gene	/db_xref="taxon:3702"	
	/chromosome="4"	
CDS	/clone="U10153"	
	/note="This clone is in pUNI 51."	
3'UTR	ecotype: Columbia"	
	1..1516	
BASE COUNT	/gene="At4g32940"	
	1..1485	
ORIGIN	/codon_start=1	
	/evidence=experimental	
Query Match	/product="putative vacuolar processing enzyme gamma-VPE"	
	/protein_id="AA15210.1"	
Best Local Similarity	/db_xref="GI:16323432"	
	/translation="MATMTVRVGVVFLVLLVSLVAVSARSPPDDVVKLPQASRF	
Matches	FRANDDSNSCTRAVLVAGSGYWNRYHOADICHAYOLLRKGLKENIVFYFD	
	DIANNENPRFIIISPHGKNDVQVPRKIDGVDVNDLFAVLIDKDTAVKGGSGK	
29.28; Score 454.8; DB 8; Length 1516;	VDSGPNDFHIFYSOHGPGVILMPTSPYLIANDLNDLKRHALGTYSKLVFLYEA	
	CESGIFEGLLPELNLNATTNAEISSWGTYPGEESPPEPEYETCLGDLYSVAMM	
60.58; Pred. No. 8.3e-93;	EDSGMNLQETLHQOYELVYKRRTPVGYSGYSHVMQYGDVIGSKNDLDMGTNPAN	
	DNFTFADANSLKILFISGPELVNKRVSAGLOPLVDDNKLNOVRAFERHCGSLSQYGI	
0; Mismatches 507; Indels 6; Gaps 2;	KMRSFANICNAGIOMQMEEAASQACTTLTPGFWSLNRGFSN"	
	1486..1516	
418 a 316 c 368 g 414 t	/gene="At4g32940"	
	1..1516	
141 TGTGGAATCTAGGCACAGAGTGCACACAGTGGCTGTTTGTAGTGTGCTGATC 200		
	141 TGTGGAACAGCAGCAGCTTACTCGGTACTAGTGGCTGTTCTAGTCGCGGATC 200	
201 AATGATATATTAATCTACAGGATCAGGCTGACATATGCCGCGTATCAGATACCTCG 260		
	201 TAGCGGATATTTGGAATTTACAGGATCAGGCTGATATATGCCATGCCCTATCAACTTCTGAG 260	
261 AARAGCGGTTTAAAGATGAAACATCATTTGTTTATGATGATGATCGGTTTC 320		
	261 GAAAGTGGATGGAAGAGAGAGATATTTGGTATTTCATGATGATGATATTCCTAACAA 320	
321 CTGCGAGATCTTAGGCTGGAGTTATCATTAATAAACACAGATGGAGAGATGTTTATAA 380		
	321 TTACGAGATCCCAAGCGCTCGACCATTTACACAGCCCTCATCGGAAAGATGCTATCA 380	
381 AGAGTTCTTACAGGATCTACTAAAGAGCTGTGAATGTTTCAAACTTCTACAACTGTGT 440		
	381 AGGAGTTCCCAAGGATATATCTGAGATGATGTCAATGTTGATAATCTATTTCGCTGTGAT 440	

RESULT 11
AF424619
LOCUS
DEFINITION

Arabidopsis thaliana AT4g32940/P2621_60 mRNA linear PLN 08-OCT-2001
1760 bp mRNA complete cds.

ACCESSION	AF2424619	AP2424619	KHMRSFANICNAGIOMBOMBEAASQACTTTLTPGPWSSLNRGFS*
VERSION	AF2424619.1	GI:15983488	1558..1760
KEYWORDS	FULL CDNA.		495 a 359 c 405 g 501 t
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis. 1 (bases 1 to 1760)		
AUTHORS	Shinn, P., Bowser, L., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Banth, J., Shinn, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.		
TITLE	Arabidopsis cDNA clones		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1760)		
TITLE	Shinn, P., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Banth, J., Bowser, L., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Banth, J., Shinn, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.		
JOURNAL	Direct Submission		
COMMENT	Submitted (26-SEP-2001) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Source		
5'UTR	Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.		
CDS	Location/Qualifiers		
	1..1760		
	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
	/clone="RAFL99-01-B01(R14418)"		
	/note="ecotype: Columbia"		
	1..72		
	73..1557		
	/note="gamma-VPE (vacuolar processing enzyme)"		
	/codon_start=1		
	/product="AT4g32940/F26P21_60"		
	/protein_id="AAL11612.1"		
	/db_xref="GI:15983488"		
	/translation="MATTWTRVSGVGLVLLVSLVAVSAARSGPDDVILKPSQASRFPAENDDDNSGTRWAVLVAGSYGVNRYHOADICHAYOLLRRGGGLKEENIVFPMYDVIANNVNPRTPTIINSPGKDVGVKPYDGDVNDVNLFAVILGDKTAVKGGSGKDVDSGNPDHIFTFYSDHSGPGVGLMPTSPYILANDNLVKKHALCTYKSLVYLEA		
	CEGSGIFEGGLPGLNIVATTASNAESSMGTGCPGPEPPPEYETCLGDLVSVAM		
	EDSGMNLQTEFLHQYELVKRTPVGYSGVSHVMGYGVGSKNDLDMCTNAN		
	DNFTADANSLKPPSRVNRDADLVHFWKRYKAPGSAKRTKAPKQVLEAMSHLH		
	IDNSVLVGLKILFGLISRGPEVLNKRVSAGQPLVDVNDNCLNKQVRAFERHCGSLSQGI		
	141 TCCTTCAATCTAGCGAAGAGCTGCAAAAAGGCACACGATGGCTGTTTCTAGTCTCGGATC 200		
	213 TCCTCAAAAACGAGGAGGATCTTAACCTCGGTACTAGCTGGCTCTCTACTCGCCGATC 272		
	201 AAATCAATATATATACTACAGCATGAGCTGACATATGCGCCGATCAGATCTCGG 260		
	273 TAGCGGATATGGAATACAGCATGAGCTGATATATGCGCATGCGCTTCACTCTGAG 332		
	261 AAAAGCGGTTTAAAGATGAACATCATTTGTTTATGTATGATGATATCGGTTTTC 320		
	333 GAAAGTGATTTGAAGAGGAGATATTTGGTATTCATGTATGATGATATCTCTACAA 392		
	321 CTCGAGAGATCTTAGGCTCGGATTTATTAATAACAGATGGAGAGATGTTTATAA 380		
	393 TTACGAGATCCAGGCTGGAACCATTTATCAACAGCCCTCATGGAAGAGATGTTATCA 452		
	381 AGGAGTCTTAAAGACTACACTAAAGAGCTGTTTAAATGTTTCAAAACTTCTACAA 440		
	453 AGGAGTCTTCCCAAGGATTTACTGGAGATGATGTTCAATGTTGATTAATCTATTTC 512		
	441 ACTTGGAAATGAAGTGGCTGACAGGAGGAGAAATGGCAAAAGTTGTGAAAAGTCTCT 500		
	513 CTTTGGAGACAAACTGCTGTTAAAGGGGAAAGTGGGAAGGTTTGGATAGTCTGCTAA 572		
	501 TGATATATCTTCTATCTTATTTATGCTGACCATGGAGCTCTGGCTTAAATAGCGATG 560		
	573 TGATCATATCTTCTATATTTCTACAGTACCATGCTGGCTCTGGAGTCTTTCGGATC 632		
	561 TGGTCTGATGAAGTTATGCGAAAGATTTCAATGAAGTCTTGGAGAGATGCTAAGAGA 620		
	633 TTCTCTCTTACCTATATATGCAATGATCTCAATGATGCTTGAAGAAGAAACATCTT 692		
	621 AAAATACAAAGATGGTGTATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTCA 680		
	693 AACATATAAAGCTTGGTGTGTTTATCTCGAAGCTTGGCAATCTGGAAGATATCTT 752		
	681 GATTTTAAAGAAATCTCAACATATACGAGTACGCTGCTTAATCTTAAAGAGAGCAG 740		
	753 GCTTCTCTGAGGTTTGAACATCTATGCCAACAATGCAATCAACCCCAAGAAAGCAG 812		
	741 CTGGGAGTCTTCTGCTGCTGATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800		
	813 TTGGGGTACCTATTTGGCTGGAGAGAACCCAGTCTCCACCCGGAGTATGAAACTTGT 872		
	801 CGCGGATACATTTAGCATCTCTCTGGCTTGAAGAGAGTACCTTCTATGACATGAGCA 860		
	873 AGGTGACTTTGATACAGTGTGCTTGGATGGAAGATAGTGGTATGCACAAATTTACAG 932		
	861 GACTTTGGAGCAACATACCACTGTTTAAAGAGAGAGTAGGATCTGATGTAC---CAGA 917		
	933 GACTCTGCACCAAGCAATATGAACCTTGTCAAAGAGGAGCTGCACCTCTTGGGTACT 992		
	918 GACTTCTCATGTATGCCGTTTTCGGAACAGAGAACTGCTTAAAGATTTATCTTCTCT 977		
	993 TGGTCTCTCATGTCTATGCAATATGCGGATGTAGGAATTTAGCAAGGAATATCTCG 1052		
	978 CATTCGGAAGAAATCTCGAAGAGATTAATCTTCTTCCACCGAATCTCTTCTCTCA 1037		
	1053 TATGGAACAACCTGCAATGACAAATTTTACCTTTGCGGATGCGAAT---TCACTAA 1109		
	1038 CTCTAAATTTCTGGCTTGGTCAATCCCGCGGATATCTCTCTCTATACCTCCAGAGA 1097		
	1110 GCCACCTTCAAGAGTTACAAACACGCTGATGCGAGATCTTGTTCATTTTGGGAAA 1169		

Db	1163	CAGAAATGCATATAGACGACAGTGTGGAACTTTGTTGGAAAGACTTTATTTATTTGGCATTTGAAA	1222
Qy	1227	TGCTCTAAATCTCTTAATCTCCACAAGAACACAGGACAGCCCTCTTTGTAGACGATTTGGGA	1286
Db	1223	GGCTCCTGAATTAATCTGAATGCTGTAAACCGGCTGGATCCGCACCTTCTTGTATGACTGGGA	1282
Qy	1287	TTGCTTCAAGACTCTAGTATTAACTGCTTCAGAATCACTGCGGTGTCACACGGTGCATATPACGG	1346
Db	1283	CTGCCTGAAAACCACTGTAAGGACTTTTGTAGACACACATTTGTGGATCGTTGTCTCAGTATGG	1342
Qy	1347	ATTCAAGTATACAGAGCGCTTGCCAAATATCTGCAATATGCGAGTGGATCTGTAAGCAAAAC	1406
Db	1343	GATGAACACATGAGGTCTTTTGGACACATGTGCAATGTAGGAATAAAGAGGAGCAAT	1402
Qy	1407	TGTTTCAGCCCATTTGAACAAGCTTGT	1431
Db	1403	CGGTGAAGCCTTCAGCACAAAGCTTGT	1427
RESULT 14			
LOCUS	E05717	1323 bp	PAT 29-SEP-1997
DEFINITION	Asparaginilendopeptidase gene.		
ACCESSION	E05717		
VERSION	E05717.1	GI:2173904	
KEYWORDS	JP 1993276960-A/2.		
SOURCE	Canavalia ensiformis.		
ORGANISM	Canavalia ensiformis		
REFERENCE	1 (bases 1 to 1323)		
AUTHORS	Mita,M., Takeda,O., Katou,I., Ishii,S. and Abe,Y.		
TITLE	ASPARAGINYL ENDOPEPTIDASE GENE		
JOURNAL	Patent: JP 1993276960-A 2 26-OCT-1993;		
COMMENT	TAKARA SHUZO CO LTD		
	OS Canavalia ensiformis		
	PN JP 1993276960-A/2		
	PD 26-OCT-1993		
	PF 07-AUG-1992 JP 1992231602		
	PR 07-FEB-1992 JP 92P		
	PI MITA MASANORI, TAKEDA OSAMU, KATOU IKUNOSHIN, ISHII SHINICHI,		
	PC C12N15/57,C12N9/48;		
	CC strandedness: Double;		
	CC topology: Linear.		
FEATURES	Location/Qualifiers		
source	1..1323		
BASE COUNT	390 a 239 c 315 g 379 t		
ORIGIN			
Query Match	28.5%;	Score 444;	DB 6; Length 1323;
Best Local Similarity	60.5%;	Pred. No. 2.4e-90;	
Matches	767;	Conservative 0;	Mismatches 495; Indels 6; Gaps 2;
Qy	167	AAGCACACGATGGCGTGT	TTTATGCTCGATCAATGAATATATTAACACAGGCATC 226
Db	5	AGGGACGACGTTGGCGGTTTTAAATGCTGTGTTTCCAAATGGTTACTGGAAATTAACAGGCATC	64
Qy	227	AGGCTGACATATGCCAGCGCTATCAGATACTCCGAAAAGGCGTTTTAAAGATCAAAAACA	286
Db	65	AGTCTGATGTATGCCATCGCTATCAAGTGTGAGGAAGGTTGGTCTGAAAAGAGAAATA	124
Qy	287	TCATTGTGTTTATGTATGATGATATCCGTTTCTCGGAGATPCCTTAGGCCTGACCTTA	346
Db	125	TTGTTGTGTTATGTATGATGACATTCCTTTCCATAAAGAAAACCCACGCGCTGGAATCA	184
Qy	347	TCATTATAAACCATGATGGAGAGATGTTTATAAAGGAGTTCCTTAAGGACTTACACTAAG	406
Db	185	TCCTTAACAGTCCACATGGAGATCATCTTTACAAAGGAGTACCTTAAGGATTACCTTGGTA	244

DHGGPVLGMPSTPNLYANDLNDLVKKYASGVTKSLVFLYLEACESGSIPEGLPEGL
NIYATTASNAEESWYCPGEPSPPESEYETCLGDLYSVAMIEDSKHNLQETLHE
QYELVKRTAGSKSGYSGHVMFEGDIGLSKEKLVLFMGNTNPADENFVFNENSRPFS
RVTHQRADLVHFWHKYQKAPGEGSARKVEAKQVLEAMSHRLHVDNLSILLIGILLFL
EGHAVLNKVRPSGEPLVDHDDCLSKSLVRAFERHCGSLSQYGIKHMRSIANNMCNAGIOM
ROMEAMQACPTIIPSWSLDRGFSA
BASE COUNT 400 a 264 c 361 g 412 t
ORIGIN

Query Match 28.3%; Score 442.2; DB 8; Length 1437;
Best Local Similarity 59.6%; Pred. No. 6e-90;
Matches 801; Conservative 0; Mismatches 533; Indels 9; Gaps 3;

Qy	96	CTTCTCAGCTGAGTCCCGCAAAACCCAAATTGCTCGAAGCATATGATGTGAATCTAGCGA	155
Db	51	CGTTTCTGGTGACGTCATCAAACTTCCTCTCTAGCTTCTTAAGTCTTCGCCCCCAACTGA	110
Qy	156	CAAGAGTGCAAAAGGCACACAGATGGGCTGTTTGTAGTTGCTGGATCAAAATGATATATAA	215
Db	111	AAACGACGATGATTCTTACTAAGTGGGCTGTTCTCGTCGCCGGATCCAGCGGATACTGGAA	170
Qy	216	CTACAGGCATCAGCTGACATATGCCAGCGGTATCAGATATCTCCGAAAGCGGTTTAAA	275
Db	171	TTATCGTCATCAGCGGATGTTGTGTCATGCTTATCAGCTTTTGAAGAAGGTGGAGTGA	230
Qy	276	AGATCAAAACATCATTTGTTTATGTATGATGATATCGCGTTTTCCTCGGAGAACTCTAG	335
Db	231	AGAGGAGAATATTTGGTGTGTTTATGTATGATGACATTCGGAAGAAGCAGGAGATCCAA	290
Qy	336	GCCTGGAGTTATCATTAATAACCATGAGAGAGATGTTTATAAGAGGTTCTTAAGGA	395
Db	291	ACCTGGAGTTTATATCATAGCTGCTTAATGGAGAGGATGCTATGATGGAGTCCCAAGGA	350
Qy	396	CTACACTAAAGAAGCTGTTAATGTTCAAAACCTCTACAAATGTTTACTTTGCAAAATGAAG	455
Db	351	TTACACTGGAGATGAAGTTAATGTGTAACCTTATAGCTGTGATTTCTTGGAAACAAAC	410
Qy	456	TGGCGTCACAGAGAGAAATGCAAAAGTTGTGAAAGTGGTCCCTAATGATAATATCTTCAT	515
Db	411	GGCTCTTAAAGGAGAGAGTGGGAAAGTTGTAGATAGCGGTCCAAACGATCATATCTTTAT	470
Qy	516	CTATTATGCTGACCATGGAGCTCTCGCTTAATAGCGATGCCACTGGTGATGAAGTTAT	575
Db	471	ATAGTATAGTATCAGCGTGTCCGGGAGTCTCGGGATGCCAACTCTCCAAAACCTATA	530
Qy	576	GGCAAAAGATTTCAATGAAGCTTGGAGAAGATGCATGAAGAAAAAATACAAACAAGAT	635
Db	531	TGCAAAATGATCTCAATGATGTCTTGAAGAAAAATATGCTTCAGAACATATAGAGCTT	590
Qy	636	GGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTGAAGGGATTTTAAAGAAAAA	695
Db	591	GGTGTTTTATTGGAGGCTTCTGAATCTGGAAGTATTTTCAAGGTCTTTTACCAGAGGG	650
Qy	696	TCTCAACATATACGAGTACTGCTGCTAATTTCAAGAGAGAGAGCTGGGAGTTTACTG	755
Db	651	TTTAAATATTTACCGCAACACTGTCATGAAATGCGAAGAAAGTAGCTGGGTACTTACTG	710
Qy	756	TCCTGAGTCATATCTCTCTCTCTCTGAGATGGAACCTGCTCGGCGATACATTTAG	815
Db	711	TCCTGGAGAGATGCTTAGTCTCTCTCTGAGTATGAGACCTGTTTGGTGACTTATACAG	770
Qy	816	CATCTCTTGGCTTGAGGACGTACCTTCTATGACATGAGCAAGAGACTTTGGAGCAACA	875
Db	771	TGTTGCTTGGATAGAAGATAGTGAACAAACAAATTTACAAACAGAGACTTTGCGACGACA	830
Qy	876	ATACCACGTTGT --- AAGAGAAAGTAGGATCTGATGTACACAGACTTCTCATGTATG	932
Db	831	ATATGAATTTGGTGAAGAGAAAGAGACTGCGAGTTCTGGTAAAGTCTTATGGTCTCATGTTAT	890
Qy	933	CCGTTTCGGAACAGAGAGATGCTTAAAGATTTATCTTCTCTTACATTTGGAAGAAATCC	992
Db	891	GGAAATTTGGAGATATAGGACTCAGCAAGGAGAGCTTGCTCTTTTATGGGTACAAATCC	950

Search completed: May 29, 2003, 16:32:08
Job time : 4045 secs

Qy	993	TGAAAACGATAAAGTCTCACTTTTACGGAATCCCTTTCTCACCACCAATCTCTAATTTCTGGCTT	1052
Db	951	AGCAGATGAAAACCTTACCTTTTGTGTAATGAGAAAT --- TCAATAAGGGCGCTTCAAGAGT	1007
Qy	1053	GCTCAATCCGCGCATATTTCTCTGCTATACCTCCAGAGAAAGATTTCAAAAAGCTCCAAT	1112
Db	1008	TACAAACGAGCTGATCGGATCTTGTCATTTCTGGCATAAAGTATCAAAAAGGCCACCGGA	1067
Qy	1113	GGGATCACTTGAAAAGCAAAAGCTCAGAAAATTTGCTTGACGAAAAGATTCATAGGAA	1172
Db	1068	AGGTCAGCAAGAAAAGTTGAAGCTCAGAACCAAGTCTTGAAGCAATGTCTCAGAGACT	1127
Qy	1173	ACAAATCGATCAGAGCATTTACAGACATTTCTCGGCTTTTCAAGTTAAACAAACCAATGTCTT	1232
Db	1128	TCATGTTGATAATAGCATTTCTGTTGATTTGGGATTTCTTTTGT --- GGTTTGGAAGGTCA	1184
Qy	1233	AAATCTCTTAACCTCCACAAAGACAGGACAGCTCTTTCTAGACGATTTGGGATTTGCTT	1292
Db	1185	TGCGGTGTTAAATAAGTCCGGCTTCTGGAGAACCGCTTGTGACGATTTGGGACTGCTT	1244
Qy	1293	CAAGACTCTAGTTAATAGCTTCAAGAAATCACTCGGCTGCAACCGGTGCAATTTACGGATTGAA	1352
Db	1245	TAAATCTCTGCTGAGAGCTTTTCGAGAGGCACTTTGCTCAGTACGGAATAA	1304
Qy	1353	GTATACAGGAGCGCTTGCCAAATATCTGCAATATATGGGAGTGGATGTGAAGCAAACTGTTTC	1412
Db	1305	GCACATGAGGTGCGATTTGCAACATGTGCAACGAGGATTCAGATGAGGCAAAATGGAGGA	1364
Qy	1413	AGCCATTCAACAAAGCTTTCTTCCA	1435
Db	1365	GGCAGCAATGCAAGGCTTGTCOCA	1387

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: May 29, 2003, 17:10:14 ; Search time 3075 Seconds
(without alignments)
4410.376 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFOILVFLHLLIF.....CNMGVDVKQTVSAIEQAQCSM 466

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Watch 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09934066/runat_19052003_163659_24353/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09934066.ecgn_1.1.3745 @runat_19052003_163659_24353 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :
1: gb_ba : *
2: gb_hgtg : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pr : *
10: gb_ro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vi : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_mu : *
20: em_or : *
21: em_ov : *
22: em_pat : *
23: em_ph : *
24: em_pl : *
25: em_ro : *
26: em_sts : *
27: em_un : *

29: em_vi : *
30: em_hgtg_hum : *
31: em_hgtg_inv : *
32: em_hgtg_other : *
33: em_hgtg_mus : *
34: em_hgtg_pln : *
35: em_hgtg_rod : *
36: em_hgtg_mam : *
37: em_hgtg_vrt : *
38: em_sy : *
39: em_hgtgo_hum : *
40: em_hgtgo_mus : *
41: em_hgtgo_other : *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	2454	100.0	1563	8	AY120765 Arabidops
2	2446	99.7	1541	8	AY084227 Arabidops
3	2188	89.2	83253	8	AP000383 Arabidops
4	1269.5	51.7	1724	8	D89971 Vigna mungo
5	1268.5	51.7	1834	8	VSCYSPROB
6	1262.5	51.4	1839	8	CSCYSPRNS
7	1254.5	51.1	1576	8	AF238384
8	1252	51.0	1859	8	AF238384
9	1246	50.8	1485	8	D89972 Vigna mungo
10	1246	50.8	1516	8	AY133531 Arabidops
11	1246	50.8	1760	8	AF059104 Arabidops
12	1246	50.8	1779	8	AF424619 Arabidops
13	1243.5	50.7	1850	8	AF370160 Arabidops
14	1236	50.4	1980	8	PV299956 Phaseolus v
15	1230.5	50.1	1437	8	BV0309173 Beta vulg
16	1230.5	50.1	1616	8	AY090296 Arabidops
17	1228.5	50.1	1866	8	AB025310 Oryza sat
18	1220	49.7	1323	6	E05717 Asparaginil
19	1219	49.7	1572	8	AF260827
20	1215	49.5	1323	6	E05718 Asparaginil
21	1208.5	49.2	1919	8	AF082347 Zea mays
22	1191.5	48.6	1864	8	ZMA131719 Zea mays
23	1183.5	48.2	1569	8	ZMA131718 Zea mays
24	1166.5	47.5	1529	8	LES243876 Lycopersi
25	1160	47.3	1461	8	AY059156 Arabidops
26	1160	47.3	1643	8	AF367254 Arabidops
27	1118.5	45.6	1649	8	AF169973 Sesamum i
28	1115.5	45.5	1652	8	PV299957 Phaseolus v
29	1109	45.2	1762	8	RCCVPE D17401 Castor bean
30	1094	44.6	1140	8	VNA238882 Vicia nar
31	1092.5	44.5	1642	8	AY062178 Oryza sat
32	1080	44.0	1323	6	E05716 Asparaginil
33	1080	44.0	1323	6	E05720 Asparaginil
34	1080	44.0	1836	8	CENAEI D31787 Jack Bean m
35	1080	44.0	1840	8	VNCYSTPRO 299174 Vicia narbo
36	1072.5	43.7	1437	8	NTA238880 Nicotiana
37	1071	43.6	1736	8	D28876 Soybean mrn
38	1054	43.0	1851	8	VSA007743 Vicia sat
39	1051.5	42.8	1360	8	NTA238881 Nicotiana
40	1051	42.8	1485	8	AF082346 Hordeum v
41	1046.5	42.6	1715	8	AF169019 Glycine m
42	1035	42.2	45160	8	AC130727 Oryza sat
43	1035	42.2	143379	2	AC130728 Oryza sat
44	1029	41.9	656	8	AY070375 Arabidops
45	977	39.8	1152	6	E05721 Asparaginil

ALIGNMENTS

RESULT 1

AY120765 1563 bp mRNA linear PLN 22-JUN-2002
 LOCUS Arabidopsis thaliana vacuolar processing enzyme/asparaginyl
 DEFINITION endopeptidase, putative (At3g20210) mRNA, complete cds.
 ACCESSION AY120765
 VERSION AY120765.1 GI:21539540
 KEYWORDS FLI-CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1563)
 Southwick, A., Karlin-Neumann, G., Nguyen, M., Tripp, M., Miranda, M.,
 Palm, C. J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M. K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S. X., Narusaka, M., Pham, P. K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R. W.
 Direct Submission
 Submitted (11-JUN-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN
 Arabidopsis Full-length cDNA"); Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Southwick, A.,
 Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C. J.,
 Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M. K.,
 Kim, C., Lin, J., Liu, S. X., Pham, P. K., Sakano, H., Shinn, P.,
 Yamada, K., Ecker, J., Theologis, A. and Davis, R. W.
 Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed
 equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R. W.
 (SSP/Stanford) contributed equally to this work as PIs.
 Location/Qualifiers
 1..1563
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL09-06-A19 (R16622)"
 /note="This clone is in pBluescript
 ecotype: Columbia"
 1..1563
 /gene="At3g20210"
 /note="MAL21.23"
 63..1463
 /gene="At3g20210"
 /codon_start=1
 /product="vacuolar processing enzyme/asparaginyl
 endopeptidase, putative"
 /protein_id="AA053323.1"
 /db_xref="GI:21539541"
 /translation="MSSPLGHFQILVFLHALLIFSAESRKTQLLNDVSSDKSAG
 TRWVLAGSNEYNYRQADCHAYQILRKGGLKDENIIVFMDIAFSENPRPGV
 IINKPDGDDVYKPKQYKEAVNQVFNVLNGESGVTGKNGKVKVSGPNDFIY
 YADKPGALIMPTGDEYMAKDFNEVKMKRKYKMKVIYVEACESSGMFEGILK
 NLNIVATNANSKESGVCYCPSPPPPEIGTCLDGTFSISWLEDSDLDHMSKETL
 EQQHVARRVGSVPPTSHVCFREGTLMKLDYSSYIGNPNDNFTTFESFSPIS
 NSGLVNPRIPELLQRIQKAPMSLESKAQKLLDEKNHRKQIDQSDITDLRLSV
 KOTNVLNLTSTRTTGGPFLVDDMDFCKTLVNSFKNHCATVHYGLKTYGTALANICNG
 VDVKQTVSAIQACSM"
 BASE COUNT 508 a 290 c 322 g 443 t
 ORIGIN
 Alignment Scores: 3.02e-208 Length: 1563
 Pred. No.:

Score: 2454.00 Matches: 466
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0
 US-09-934-066-2 (1-466) x AY120765 (1-1563)
 QY 1 MetSerSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuLeuPhe 20
 Db 63 ATGTCATCTCTCTGGTCACTTTTCAGATTCTTGTGTTTTCTTCATGCTTTGCTTATCTTC 122
 QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
 Db 123 TCAGCTGAGTCCCGCAAAACCAATTCCTGAACGATAATGATGTTGAATCTACGCAAG 182
 QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 Db 183 AGTGCAAAAGCACACGATGGCTGTTTATAGTGTGATCAATCAATATATATACTAC 242
 QY 61 ArgHisGlnAlaAspPheCysHisAlaTyrClnIleLeuArgLysGlyGlyLeuLysAsp 80
 Db 243 AGGCATCAGCTGACATATGCCACGCTATCAGATACCTCCGAAAGGCGTTTAAAGAT 302
 QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
 Db 303 GAAACATCATTTGTTTATGATGATATCCGTTTTCTTCGGAGAAATCCTAGCCCT 362
 QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
 Db 363 GGAGTTATCATTAATAACCCAGATGGAGAGATGTTTATAAGAGAGTTCCTAAGGACTAC 422
 QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
 Db 423 ACTAAGAAGCTGTTAATGTTCAAAACCTTCTACAAATGTTTACTTGGAAATGAAAGTGC 482
 QY 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
 Db 483 GTCACAGAGGAAATGGCAAGTTGTGAAAAGTGGTCTTAATGATATATATCTTTCATCTAT 542
 QY 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
 Db 543 TATGCTGACCATGGAGCTCCTGGCTTAATAGCGATGCCACTGGTGATGAAGTTATGGCA 602
 QY 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
 Db 603 AAAGATTTCAATGAAGTCTTGGACAACATGCATAAGAGAGAAAAATACACAAGATGGTG 662
 QY 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
 Db 663 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 722
 QY 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
 Db 723 AACATATACCGAGTCACTGCTGCTAATTTAAAGAGAGACGCTGGGGAGTTTACTGTCTCT 782
 QY 241 GluSerTyrProProProProSerGluIleGlyPheCysLeuGlyAspThrPheSerIle 260
 Db 783 GAGTCATATCT 842
 QY 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
 Db 843 TCTTTGGCTTGAGGACAGTGAACCTTCATGACATGAGCAAGAGAGACTTTGGAGCAACAATAC 902
 QY 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
 Db 903 CACGTTGTAAGAGAGAGAGTATGATCTGATCTACCGAGACTTCTCATGTATGCCCTTTC 962
 QY 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
 Db 963 CGAACAGAGAGAGTCTTAAAGATTATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022
 QY 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340

Db 1023 GATAACTTCACTTTCAGGAATCCTTTTCTCACCATCTCTAATTCGGCTTGGTCAAT 1082
 Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
 Db 1083 CCGCGGATATCTCTCTGCTATACCTCCAGAGAAGATTCAAAAGCTCCCAATGGGATCA 1142
 Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
 Db 1143 CTTGAAGCAAGAAGCTCAGAGAAATTCCTTGACGAAAGAATCATAGAAACAATC 1202
 Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
 Db 1203 GATCAGAGCATTCAGACATCTTCGGCTTTCAGTTAAACAACCAATGCTTAAATCTC 1262
 Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
 Db 1263 TTAACCTCCACAGAACACAGACAGCCCTCTGTAGACGATGGGATGCTTCAAGACT 1322
 Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
 Db 1323 CTAGTTAATAGCTCAAGATCACTGCGGTGCAACGGTGCAATACGGATTGAAGTATACA 1382
 Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
 Db 1383 GGAGCGCTTCCCATATCTGCAATATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1442
 Qy 461 GluGlnAlaCysSerMet 466
 Db 1443 GAACAAGCTTGTTCGATG 1460

RESULT 2
 LOCUS AY084227 1541 bp. mRNA linear PLN 21-JUN-2002
 DEFINITION Arabidopsis thaliana clone 100372 mRNA, complete sequence.
 ACCESSION AY084227
 VERSION AY084227.1 GI:21402937
 KEYWORDS FLU_CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1541)
 Haas,B.J., Volkovsky,N., Town,C.D., Troughan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
 Full-length messenger RNA sequences greatly improve genome annotation
 Genome Biol. (2002) In press
 2 (bases 1 to 1541)
 Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
 Full-length cDNA from Arabidopsis thaliana
 Unpublished
 3 (bases 1 to 1541)
 Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
 Direct Submission
 Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
 COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0.

Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES
 source
 1..1541
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="100372"
 59..1459
 /codon_start=1
 /product="vacuolar processing enzyme/asparaginyl
 endopeptidase, putative"
 /protein_id="AA060827.1"
 /db_xref="GI:21536495"
 /translation="MSSPLGHFOILVFLHALIFSAESRRKTQLLNDNDVSSDKSAK
 TRWALVAGSYEYNYRHQADICHAYILRKGLKDENIIVFMYDDIARSSENPRPGV
 IINKPGEDVYKGVKPKDYTKAVNVONFYNVLLGNESGVTKGNGKVKVSGPNDFIY
 YADHGAPGLIAMPITGDEVAKDFNEVLEKHKRKLIYKMWIYVEACESGSMFGILAK
 NLNIYAVTAANSKESWGVCYPPPEIGTCIGDTGTFISIWLSDSDUHDMSKEPL
 EQQYHVVKRRVSDVPETSHVCRFGTEKMLKDYLSYIGRNPNDFNFTFESSPILS
 NSGLVNPDIPLLYLQKKIQAPMGSLEKQAKLIDENKHKRQIDQSTIDILRSV
 KOTNLNLILTSTRTTGTGQPLVDDWDCFKTLVNSFKHCGATVHYGLKYTGALANICNG
 VDYKQTVSAIEQACSM"

BASE COUNT 499 a 287 c 321 g 434 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1..52e-207 Length: 1541
 Score: 2446.00 Matches: 465
 Percent Similarity: 99.79% Conservative: 0
 Best Local Similarity: 99.79% Mismatches: 1
 Query Match: 99.67% Indels: 0
 DB: 8 Gaps: 0

US-09-934-066-2 (1-466) x AY084227 (1-1541)

Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
 Db 59 ATGTCTAGTCTCTCTGTCACCTTCAGATTCCTGTTTTCTTCATGCTTATCTTC 118
 Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
 Db 119 TCAGCTGAGTCCGCAAAACCAATTGCTGAACGATAATGATGTTGAATCTAGGCAAG 178
 Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 Db 179 AGTCAAAAGCACACGATGGCTGTTTAGTCTGGATCAATGATATATACTAC 238
 Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAsp 80
 Db 239 AGGCATCAGGCTGACATATGCCACGCGTATCAGATACCGAAAAGCGGTTTAAAAGAT 298
 Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
 Db 299 GAAACATCATPTGTGTTTATGATGATGATCGCTGTTTCCGCGAAGATCCTTAGGCCT 358
 Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
 Db 359 GGAGTTATCATTAATAAACCCAGATGGAGAAGATGTTTATAAGGAGTTCCTAAGGACTAC 418
 Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGluAsnGlyGly 140
 Db 419 ACTAAAGAGCTGTATATGTTTAAACACTTCTACAAATGTTTACTTGGAAATGAAAGTGC 478
 Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerClyProAsnAspAsnIlePheIleTyr 160
 Db 479 GTCACAGGAGGAATGGCAAAAGTTGTGAAAAGTGGTCTTAATGATAATATCTTCATCTAT 538
 Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
 Db 539 TATGCTGACCATGGAGCTCTCGCTTAATAGCATGCCACTGCTGATGAAGTATGCA 598
 Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200

```

Db      599  AAGATTTCAATGAAGTCTTGAGAGATGCATAAGAGAAAATATACAAAGATGGTG 658
Qy      201  ILeTYrValGlUAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db      659  ATCTATGTTGAAGCATGTGAATCAGGAGATGTTTCAAGGGATTTTAAAGAAAATCTC 718
Qy      221  AsnIleTYrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTYrCysPro 240
Db      719  ACATATACCGAGTCTGCTGCTTCTAGATTCATTAAGAGAGACGCTGGGAGTTTACTGCT 778
Qy      241  GluSerTYrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db      779  GAGTCATATCCCTCCCTCTGAGATGGAACTTCTCGCGGATACATTAAGCATC 838
Qy      261  SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnInTYr 280
Db      839  TCTTGCTTGAGGACAGTGCCTTCATGACATGACCAAGAGACTTTGGAGCAACAATAC 898
Qy      281  HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db      899  CAGCTTGTAAAGAGAGTAGGATCTGATCTACAGAGACTTCTCATGTATGCCGTTTC 958
Qy      301  GlyThrGluLysMetLeuLysAspTYrLeuSerSerTYrIleGlyATGAsnProGluAsn 320
Db      959  GGAACAGAGAAATGCTTAAGATTAATCTTCTTACATTCGGAAGAAATCTGAAAC 1018
Qy      321  AspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsn 340
Db      1019  GATACTTCACCTTCACCGAATCTTTCTCACAATCTCTAATTCGCTGGCTGGTCAAT 1078
Qy      341  ProArgAspIleProLeuLeuTYrLeuGlnArgIleGlnLysAlaProMetGlySer 360
Db      1079  CCGCGGGATATCTCTCTATACCTCAGAGAAAGATTCAAAAGCTCCAAATGGGATCA 1138
Qy      361  LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db      1139  CTGTAAGCAAAAGAGCTCAGAAGAAATGCTTCACGAAAAGATCATAGAAACAATC 1198
Qy      381  AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db      1199  GATCAGAGCATTTACAGACATTTCTGGCGCTTTCAGTTAAACAACAATGCTTAAATCTC 1258
Qy      401  LeuThrSerArgThrArgThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db      1259  TTAACCTCCACAAGAACACAGGACAGCTCTTGTAGACGATTGGGATTCCTCAAGACT 1318
Qy      421  LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTYrGlyLeuLysTYrThr 440
Db      1319  CTAGTTAATAGCTTCAAGAACTACCTCGCTGCACAGCTGCTTACGAGATTGAAGTATACA 1378
Qy      441  GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db      1379  GGAGCCCTTGGCAATATCTGCAATATGGAGTGGATGGAAGCAAACTGTTTCAGCCATT 1438
Qy      461  GluGlnAlaCysSerMet 466
Db      1439  GAACAGCTTGCTCGATG 1456

RESULT 3
AP000383
LOCUS
DEFINITION
AP000383 Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone: MAL21.
ACCESSION
VERSION
AP000383.1 GI:5672520
KEYWORDS
SOURCE
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MAL21
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eucotyledons; Magnoliophyta; eudicotyledons; core eudicots;
Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (sites)

```

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Kaneko, T., Katoh, T., Sato, S., Nakamura, A., Asamizu, E. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 3. II.
Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
DNA Res. 7 (3), 217-221 (2000)
20363099
2 (bases 1 to 83253)
Kaneko, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E. and Tabata, S.
Direct Submission
Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MAL21
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Heagard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MZE19 and the 3' clone is MOC12.
Location/Qualifiers
1. 83253
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MAL21"
/clone_lib="Mitsui P1"
complement(1..376)
/product="nucleoid chloroplast DNA-binding protein-like"
/note="CDS is reported in Acc# AP002050
gene_id:MZE19.7"
/number=1
/evidence=not_experimental
complement(join(4273..4315,4392..4510,4727..4902,
5273..5377,5836..5914,6019..6096,6222..6290,6555..6620,
7231..7266,7418..7486,7659..7724,7870..8163))
/note="gene_id:MAL21.2"
/codon_start=1
/evidence=not_experimental
/product="protein arginine N-methyltransferase-like
protein"
/protein_id="BAB01859.1"
/db_xref="GI:9293956"
/translation="MQSGDFNSGPHGDHRELEDKQPSLSFSGRAKKRSHAGAR
DPRGLANLVRSVDLGEHKSLTESPPCTDFDVAVFHSAHVHHEMIKDRAR
TETRYREALMOHOSLIEGKVVVDVGGCTGTLSIFCAOAGAKRYAVDASDIAVOAKEVY
KANGLSKDVILVHGRVDEIDEVDVVIIESEMMGYMLLYESNMGVITARDMLKPG
LILPSHATLYMAPISHPDYRSHSIDFWRNVYGDMSAMMOLAKOCAPPEPSISCE
NVLTPVFPPLFSGFPASSPAKNTSETSIASGSSISPSGVEVQKRTNPDSALVLS
TPSPSPHTHWOOTIVFYDIDVEDQDVIEGVSVTLSQSKENKRMNTHLEYSSAGRSF
VKESVNR"
10472..11680
/note="emblicAB78723.1
gene_id:MAL21.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01860.1"
/db_xref="GI:9293957"

```

/translation="MTMMSDLSDLEELISRVPRSTSLGAVRSTCKRWNTLFKDRILC
KAERTDQRFIMKKYKLCVSRFDNLGNTMEDGTEFVDPISIKELGHFFNQVTKISVF
QDGLLLCVRILVWVNPYLGQIRWIESGNTNRLFDVAIGYDNNRKHKILRF
LEYVNNKMRLELEYEDFSSNSWRLDIAPWEIESYORGASLKGNTYFIAKEKIE
YEEDGEPADNLLCFDFTTESGFQFLPFPQHYLDVGALSGLGDEKLAALFOQGD
TDUSEIEMWTLTETNTVSNWNPFLKVDMEPHYRGSFMDYIGSGFIDEKELAVGF
QDFSGREYEDATYIIGENGYKKVRLGAPANOQGCFFSPVCFSSYVPSLVQINOI
VGFKKEREET"
join(15535..15806,16005..16155,16253..16435,16674..16748,
16945..17100,17229..17327,17425..17552,17741..17819,
17917..18282)
CDS
/note="gene_id:MAL21.4"
/codon_start=1
/evidence=not_experimental
/product="chloroplast outer envelope hexokinase"
/protein_id="BAB01861.1"
/db_xref="GI:9293958"
/translation="MGKVLVMTAAAVACSVATVVRMRKGRKRWRRVGLLKDL
EACETPLGRKQWDAIAVEMOAGLVSEGSKLMLLTVDLDPNGSETGTYYALHL
GGSYFIRIKVHLGQSRSLQVQVERHSIPTSLMNSTSEVLPDFLASSLQRTIEKEN
DFSLOPLKRELAFTEFPVKQTSISSGVLTKWTFKPAISEMAGEDAECLOQALNKR
GLDIRAALVNDTVGALSGFHFDPDTIAAVFGTSGNACVLERTDAIKCONPRTTS
GSMVNMENGFSSRLPRTSYDLELDAESMNSNDMGFEKMGIMGLGDIVRVILRM
SOESDIFGPSSILSTFVLRTNSVSNHEDDISELQEVARIKDLGSEVPWKRKL
VVKIDVVTARRAALAAAGIAGILKKVGRDGGGRRSDKOIMRRTVVAVEGGLYLN
RMFREYDEALRDLGEDVAQHVVVVVKAMEDGSSIGSALLASSQSQTIPSV"
complement(join(18762..18845,18933..19038,19118..19221,
19306..19434,19520..19624,19780..19914,19994..20080,
20208..20304,20550..20767,20863..20928,21116..21211,
21364..21414,21501..21560,21669..21755,21869..21957,
22170..22234,22412..22484))
/note="gene_id:MAL21.5"
/codon_start=1
/evidence=not_experimental
/product="chaperonin, t-complex protein alpha subunit"
/protein_id="BAB01862.1"
/db_xref="GI:9293959"
/translation="MSTSAQNPDISGRQSDVVRTQNVMACQAVSNIVKTSIGPVGL
DKMLDDIGVITINDGATILRMEVEHPAAKVLVELAEQDREVGGDTTSVIVAAE
LLKRLANDVRNKTHPTSIISGYRLAMRECKYIEEKLVTKVEKLGKVPILNCAKTSWS
SKLISGDSDFANLVAVSVKVTNORGEIKYPIKGINILKAHGOSARDQSYLLNGA
LNTGRAAGWPLVSPAKIACLOFNQKTKMQLGVQVVDNPDELEKIRQREADMTKE
RIEKLKAGANVILTKGIDDMALKYFEAGALARRVRKEDMRHAKATGATLVTFE
ADMEGETFPAHUGSADVEVERIADDDVILKGTSTSAVSLIRGANDMLDNE
RALHDLICVYKRLSTNVAGGAVESALSVEHLIATLIGSREQIAIEAFADALI
IKPVLYNAAKDTELAKRLRAYHTTAQTKADKKHYSKRKRKCSMGLDLVNGTIRNLE
AGVTEPAMSVKIIQFATEAATILRIDDMIKLVKDESGEE"
complement(join(23145..23191,23289..23348,23451..23514,
23623..23850,23934..23961,24383..24501))
/note="gene_id:MAL21.6"
/codon_start=1
/evidence=not_experimental
/product="ubiquitin conjugating protein-like"
/protein_id="BAB01863.1"
/db_xref="GI:9293960"
/translation="MATVNGYTGNTPAATTPAATGSKQSPAPPTKTVDSHVLKRLOSE
LMGLMGADGIFSAFPEEDNIFCWKGTITGSKDTVEFGETYRLSLTFSDNYPFKSPKV
KFETCCFHPNVVDIGNCLDLODKWSSAYDVRTILLISQSLGEPNISSPLNQNAQ
LWNSOEYRKWVEKLYPLNA"
complement(join(25168..25258,25476..25644,25764..25981,
26173..26234,26335..26425,26620..26689,26796..26843))
/note="gene_id:MAL21.7"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01864.1"
/db_xref="GI:9293961"
/translation="MEALYAKLYDKYTKLQKKYSEYDEINKQEKEKFLTFVSASEEL
MEHLRGENSLSEWEKRLNEIISIRGRDDKFLQCKLLMEELKXNLSLESEVVKIK
ELVQEHPRNVDOQSGKKRKPTPEARVTRSMIKRSLSEDLVETDWMVSPDIXKH
KAKEPLVPOCCRTYDGSSSASCTFOALGKHLGLMKLSTNNKGRACIVASHT
TGLSFLSTFTNNPGESESLYPASLGTFQRPVAPEMREVIFKSTSMCFIFFEVSR
VILKNC"

```

```

27734..27973
/note="contains similarity to cytochrome P450"
/gene_id:MAL21.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01865.1"
/db_xref="GI:9293962"
/translation="MMLQCEDWTKTGDKVDMEEAIRGLNLTMAHLKCIPIVARNLNL
ACNLKPCNEFAKLLVRGSSAPLCMFPCPSLPV"
join(29031..29951,30031..30681)
/note="gene_id:MAL21.9"
/codon_start=1
/evidence=not_experimental
/product="cytochrome P450-like protein"
/protein_id="BAB01866.1"
/db_xref="GI:9293963"
/translation="MAAMNIVDFQNCFFVLLSLFSLCYLSLFRKPKERLHYDLPP
SPPTPIIGHLLLSVLVHRSIQTSKYSGILYLRVFSFVALVSASAIYEIFRE
HDNISFRPPPTDDSLFAGSFSTSAFYDGMFKMKLLVTLNLGPOALERSGFRA
DELDFYENLLDKAMKESVDICVEALKLSNNSICKIMGRSCSEENGEAEVRALAT
QDLDTKALLANNLRAGFKLVLSLFRKEMDVSRFDELLERILVHEHDKLDMHHQ
GDLVDALLAACRDKNAEYKISRNIHKSFFADLLFASTDTFVOTTKTVAELINPNV
LERLCRIDSVVGKARLIQETDLPNLPLQAVYKEGLRHPGGLPFAFESQEGCRIGG
FYVPEKTLMINAYVMRSDSWEDPDEFKPERFLASSRSEQEKERREQALKYIATGS
GRSCPGTENLAYIFLGTAGVMVQGFEMRIKEEKNMEEANVGLSLTMAYPLKVTVP
RTLVPITQNPLNRSS"
join(32158..32475,32844..33086,33190..33813)
/note="contains similarity to cytochrome P450"
/gene_id:MAL21.10"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01867.1"
/db_xref="GI:9293964"
/translation="MAAMINVEFQNCFFIMLLCFLCYLSLLFRKPSRONSVDCLL
PPSPSPPIIGHLLLSVLVHRSIQTSKYSGILYLRVFSFVALVSASAIYEIFRE
RVHDKFFLAATLRRLEKLGILSUFKEIMGVSRGFDLELLERILREHEEKEDDEHDTOM
MDALLAANDYNDKAEYKTRNQIKAFIVDMFIAGTDSALTQGTMAEIIINNPNVLR
REKIDSVVGSRILQETDLPKLPYSAQVYKEGLRHPPTPLMVREFQEGKVGIFYIP
ASTTLVVGAVAMRDPNVWEEPEEFKPERFLASSRSEEEIREQALKYIATGSGRGG
CPGTNIAYIFVGTAGIMVMVQCFDKWIKGDKVDMKEALGGLNLTMLAHLKCTPVARS
PF"
join(34864..34912,34957..35241,35310..35443,35870..36133)

```

Alignment Scores:

Pred. No.:	2,32e-182	Length:	83253
Score:	2188.00	Matches:	463
Percent Similarity:	67.89%	Conservative:	0
Best Local Similarity:	67.89%	Mismatches:	3
Query Match:	89.16%	Indels:	219
DB:	8	Gaps:	7

US-09-934-066-2 (1-466) x AP000383 (1-83253)

Qy	1	MetsSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20
Db	72700	ATGCTAGTCCTCTGTGTCACCTTTTCAGATCTTTCTTCATGCTTTGCTTATCTTC	72759
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys	40
Db	72760	TCAGCTGAGTCCCAAAACCCCAATTCCTGAACGATTAATGATTTAGTACGACAG	72819
Qy	41	SerAlaLysGlyThrArgTyrAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr	60
Db	72820	AGTCAAAAGGCACACGCTGGCTGTTTATGTTGCTGATCAATGAATATATACTAC	72879
Qy	61	ArgHis-----	62
Db	72880	AGGCATCAGTTGTGTTAAATTATGTTTGAAGCTTTAACAACAAAAAGGTCCTA	72939
Qy	63	-----GlnAlaAspIleCysHis	68
Db	72940	AGCGAGATTTGTATGAACATAATCGACCGACGCTTTTATTTTCACAGGCTGACATATGCCAC	72999

QY 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleValPheMetTyr 88
 DB 73000 GCGTATCAGATACCTCGAAGAGCGGTTTAAAGATGAAACATCATTTGTATGTAT 73059
 QY 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
 DB 73060 GATGATATCCGTTTCTTCGGAGATCTAGCCCTGGAGTTATCATTAATAAACCCAGAT 73119
 QY 109 GlyGluAspValTyrLysGlyValProLys----- 118
 DB 73120 GGAGAAGATGTTTATAAAGAGAGTTCTTAA- GGTCTCTATTCTACTTCTTTTGGCGTTA 73178
 QY 118 ----- 118
 DB 73179 TTTCTACGTTGAATTCAATATACATATATATTCAAGTTTGTGTTATGTTGGGTAG 73238
 QY 119 AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu 138
 DB 73239 GACTACACTAAAGAGCTGTTAATGTTCAAACTTCTACAATGTTTACTTTGGAAATGAA 73298
 QY 139 SerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe 158
 DB 73299 AGTGGCGTCAAGAGGAGAACTGCGAAAGTGTCTTAATGATATATCTTC 73358
 QY 159 IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle----- 170
 DB 73359 ATCTATTATGCTGACCATGGAGCTCTGCTTATAGTTTCTTAATTTTATGAATTA 73418
 QY 170 ----- 170
 DB 73419 TTACGTACCATCAATCCATATCTATATAAAGATTTTCTTGATACACTAGCAACCGCG 73478
 QY 171 -----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu 187
 DB 73479 ATTTTCTCAGCGATGCCACCTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTG 73538
 QY 188 GluLysMetHisLysArgLysLysTyrAsnLysMetVal----- 200
 DB 73539 GAGAAGATGCATAAGAGAAAAATACAAAGATGGT-ATATAACTCAACCATTCGTTA 73597
 QY 200 ----- 200
 DB 73598 CCTAGCTTTATACATATGTGTTCTGTTTGTGAATCTCTATGGTGTTGTTTGGATGTT 73657
 QY 201 -----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218
 DB 73658 TAGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTCAAGGGATTTTAAAGAA 73717
 QY 219 AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyr 238
 DB 73718 ARTCTCAACATATACGAGTGTGCTGCTAAATCTTAAAGAGAGAGCTGGGGAGTTTAC 73777
 QY 239 CysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPhe 258
 DB 73778 TGCTCGATGATATCTCTCTCTCTCTGAGATTGGAACCTGTCTCGCGGATACATTT 73837
 QY 259 SerIleSerTyrLeuIle----- 264
 DB 73838 AGCATCTTGGCTTGAGGACAGGTACTGCAACAAAAAGATTCAATCTTATGGACTA 73897
 QY 264 ----- 264
 DB 73898 TTCGAATGATTGTTGTTGGAAGATATTTGTTTCTATGTTTCTGTTGTT 73957
 QY 265 -----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnIntyrHisVa 282
 DB 73958 GTTTGGGACAGTACCTTCATGATGATGAGCAAGAGACTTTGGAGCAACAAATACACGT 74017
 QY 282 lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyTh 302
 DB 74018 TGTAAAGAGAAGTAGGATCTGATCTACAGACATCTCATGTATGCGGCTTCGGAAC 74077
 QY 302 rGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAs 322

DB 74078 AGAGAAGATGCTTAAGGATTTCTTCTCTACATTGGAAGAAATCCTGTAACCATATAA 74137
 QY 322 nPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsnProAr 342
 DB 74138 CTTCACTTTCACGNAATCCTTTCTCTCACAATCTCTAAATCTGGCTTGGTCAATCCGG 74197
 QY 342 gAspIleProLeuLeuTyrLeuGlnArgLysIle----- 353
 DB 74198 CGATATCTCTGCTATACCTCCAGAGAAAGGT- GAGCTTTTTTTCGGGTTTTTGTATCAT 74256
 QY 354 -----GlnLysAl 356
 DB 74257 TTTAAACGAAAGAGTTTTTCAGCATGTTTAAATGTTTATTCATCTCTAGATTCAAAAAGC 74316
 QY 356 aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHI 376
 DB 74317 TCCAATGGGATCACTTCAAGACAAAGAACTCAGAAGAAATGCTTGACGAAAGAAATCA 74376
 QY 376 sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs 396
 DB 74377 TAGAAACAAATCGATCAGAGCATTCACAGACATTCGCGGCTTTCAGTTTAAACAACCAA 74436
 QY 396 nValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAs 416
 DB 74437 TGTCTTAAATCTCTTAACTTCCACAAGAACAGACAGACGCTCTGTAGACCATTTGGGA 74496
 QY 416 pCysPheLysThrLeu----- 421
 DB 74497 TTGCTTCAAGACTCTAGTAACAAACCAATCTCAAACTTGTGTTACTGTTCTACGCCAA 74556
 QY 422 -----ValAsnSerPhe 425
 DB 74557 CAACCATTTGATTTTACTTAAACCAAGTGTATATCGAATGAAATCGCAGGTTTAAATGCTT 74616
 QY 425 eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs 445
 DB 74617 CAAGAATCATCTCGGTGCAACGCTGCATTCAGGATGAAGTATACAGAGCGCTTGCCAA 74676
 QY 445 nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSe 465
 DB 74677 TATCTCAATATGGGAGTGATGTGAAGCAAACTGTTTTCAGCCATTGAACAAGCTTCTTC 74736
 QY 465 rMet 466
 DB 74737 GATG 74740
 RESULT 4
 DB89971 LOCUS
 DEFINITION Vigna mungo mRNA for asparaginyl endopeptidase (VnPE-1), complete cds
 ACCESSION D89971
 VERSION D89971.1 GI:4589395
 KEYWORDS asparaginyl endopeptidase (VnPE-1).
 SOURCE Vigna mungo seedling cotyledon cDNA to mRNA, clone:pPEA.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna
 REFERENCE 1 (bases 1 to 1724)
 AUTHORS Okamoto,T. and Minamikawa,T.
 TITLE Molecular cloning and characterization of Vigna mungo processing enzyme 1 (VnPE-1) in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP)
 JOURNAL Plant Mol. Biol. 39 (1), 63-73 (1999)
 MEDLINE 99178794
 REFERENCE 2 (bases 1 to 1724)
 AUTHORS Okamoto,T.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-1996) Takashi Okamoto, Tokyo Metropolitan

SOURCE
ORGANISM Vicia sativa
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Vicia.
1 (bases 1 to 1834)
Becker, C., Shutov, A.D., Nong, V.H., Senyuk, V.I., Jung, R.,
Horsmann, C., Fischer, J., Nielsen, N.C. and Muntz, K.
Purification, cDNA cloning and characterization of proteinase B, an
asparagine-specific endopeptidase from germinating vetch (*Vicia
sativa* L.) seeds
Eur. J. Biochem. 228 (2), 456-462 (1995)
9520376
MEDLINE
PUBMED
REFERENCE 7705362
2 (bases 1 to 1834)
Shutov, A.D., Do, N.L. and Vaintraub, I.A.
Purification and partial characterization of protease B from
germinating vetch seeds
Biokhimiia 47 (5), 814-821 (1982)
82232279
MEDLINE
PUBMED
REFERENCE 7046813
3 (bases 1 to 1834)
Shutov, A. and Vaintraub, I.
Degradation of storage proteins in germinating seeds
Phytochemistry 26, 1557-1566 (1987)
JOURNAL
TITILE
AUTHORS
REMARK (sites)
4 (bases 1 to 1834)
Becker, C.
Direct Submission
Submitted (27-JUN-1994) Becker C., Institut fuer pflanzenegenetik
und Kulturpflanzenforschung, Molecular Cell Biology, Corrensstr. 3,
Gatersleben, Sachsen-Anhalt, Germany, D-06466
Location/Qualifiers
1. .1834
/organism="Vicia sativa"
/db_xref="taxon:3908"
/clone="join: pPB14 (bases 1-1809) and pSK54 (bases
1810-1834)"
/tissue_type="cotyledon"
/clone_lib="PCR-amplified cDNA"
/dev_stage="germinating seed"
16. .1497
/function="storage protein hydrolysis"
/note="bases 164-1809 of pPB14 overlapped with pSK54,
antisense primer of pSK54 (1780-1809) was used to obtain
pPB14"
/citation=[1]
/codon_start=1
/evidence=experimental
/product="cysteine proteinase"
/protein_id="CAA84383.1"
/db_xref="GI:510358"
/db_xref="SWISS-PROT:P49044"
/translation="MGSQLSTLLEFTTIVTFLTVSSGRDLPGDYLRLEPSETSRFFR
EPKNDPFGTIRWAILLAGSNGWYRHOSDCHAYQLLRKGGKEENIVFVYDDIA
SNEENPRGVYIINKPDGDDVAGVPKDYTGAEVHADNFYAAALGNKKSALTGGSGKVD
SGPNDFIVYITDHDGPGVGLGMPGYLYASDLNEVLKHKHAGSTKSLVFLPACES
GVTFGLDLDQNIYATTASNAEISSWGYYCPGDKPPPEYSTCLGDLXSIAMNDS
EVHNQTESLOQYKLVKNRTISEPSYGVHVEYDGLSKNDLYOYLGTNPANDNSF
VDENSLKLRTPSAVNORDADLIHFWEKFKAPGEGSKNEAEKQVLEAMSRKHI
DNSVKLIGOLLGIEKGTELLDVRPAGSLVDNWDCLKTKVTKFTFCHGSLSOYGMK
HMRSFANICNAGIPNEPMAESAQACASIPANPWSLSLQGGFSA"
16. .72
sig_peptide
mat_peptide
160. .1494
/product="cysteine proteinase"
/function="storage protein hydrolysis"
/evidence=experimental
278
variation
/replace="t"
281
variation
/replace="g"

variation 528
variation /replace="a"
variation /replace="c"
variation /replace="a"
variation /replace="a"
variation /replace="t"
variation /replace="a"
variation /replace="a"
variation /replace="c"
polYA_signal 556 a 353 c 397 g 528 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 4.89e-103 Length: 1834
Score: 1268.50 Matches: 235
Percent Similarity: 69.63% Conservative: 70
Best Local Similarity: 53.65% Mismatches: 122
Query Match: 51.69% Indels: 11
DB: Gaps: 3
US-09-934-066-2 (1-466) x VSCYSPROB (1-1834)
Qy 31 AsnAspAsnValGluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeu 50
Db 157 AAGCATGACGACTTTGAG-----GGAACTAGGTGGCGCATTTTA 195
Qy 51 ValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyr 70
Db 196 CTAGCTGGTTCCTAATGGTTACTTGGAAATTATAGACATCAGTCTGATGTTGTTCATCCGCTAT 255
Qy 71 GlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAspAsp 90
Db 256 CAATTACTGAGGAAGGTGGTTCGAGGAAGAAATAATATTATTGTTTCATGTATGATGAT 315
Qy 91 IleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAspGlyGlu 110
Db 316 ATTGCTTCCAATGAAGAAATCCAAAGGCTGCTGTCTATAATTAACAAACCTGATGGGAT 375
Qy 111 AspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGluAsnPhe 130
Db 376 GATGTTTATGACGAGGACTTCCAAGGATATATCTGTCGAGAGTACATCGCGACAAATTC 435
Qy 131 TyrAsnValIleLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLys 150
Db 436 TATGCTGCTTTACTTGGAAATAAATCAGCTCTTACRGGTGGAGTGGGAAAGTTGTGGAT 495
Qy 151 SerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIle 170
Db 496 AGTGGTCCCAATGATCATATTTTGTATACATCATGATGTTGGTCCAGGGGTCTT 555
Qy 171 AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLysMet 190
Db 556 GGTATGCCCTGTTGGTCCCTTACTTGTATGCTATCTGATCTGAATGAAGTCTTCAAGAAAAG 615
Qy 191 HisLysArgLysLysTyrAsnLysMetValTleTyrValGluAlaCysGlySerGlySer 210
Db 616 CATGCTTCTGGAAACATATAAGAGCCCTAGTATTTATCTAGAGGCGATGTGAATCTGGGAGT 675
Qy 211 MetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSer 230
Db 676 ATCTTTGAAGGTCTTCTTCCAGATCATCTCAATATCTACGCGACAAAGCGTTCAATGCA 735
Qy 231 LysGluSerSerTrpGlyValTyrCysProGlyLysTyrProProProSerCysIle 250
Db 736 GAAGAAAGCAGTTGGGGATCTATTGTCCTGGGGATAAACCTCCCCACCCAGAGTAC 795
Qy 251 GlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAsp 270
Db 796 TCAACCTGTTGGGTGACCTATATCTGATGAGGAGACAGTGAAGTACACAT 855


```

Db 550 CCAGGGGTTCTCGGACGCGCTGCTGCTTACATATATGATCATCTGATTGGTTGAAGTA 609
Qy 187 LeuGluLysMetHisArgLysLysLysLysLysMetValIleTyrValGluAlaCys 206
Db 610 CTGAAGAAAACCATGCTTCTGGAACGTATATAAACCTAGTATTTTATTTGGAGGCGATGT 669
Qy 207 GluSerGlySerMetPheGluGlyLeuLysLysLysLysLysLysLeuAsnIleTyrAlaValThr 226
Db 670 GAAGCTGGAGATATCTTGAAGGTCTTCTGGAATATATATATATATATATATATATATATAT 729
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro 246
Db 730 GCTTCGAATGCGGAGAAGTAGTTGGGAAACATATCTCTGGGAGTAGTTCCTAGTCCT 789
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 790 CCCCCAGAAATATTCACCTGCTGGGTGACCTGTACAGTGTGCTGGATGGAAGACAGT 849
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 850 GATAGACAACTTCGACAGAAATCTTTACACCAACAATATAAAGTGGTTAAGGATAGG 909
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 910 ACTCTATCTGGAGGTGGTATGGCTCTCACGTGATGCGATGATGATGATGATGATGATGAT 969
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 970 AAGGATGCTCTCTCTGTTATTTGGGTACAGATCTGCAAAATGATAATTTGACTTTCGTG 1029
Qy 327 GluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeu 346
Db 1030 GATGAA--AACTCTTATGTCATCTCACTGAGTCACCAACCAAGGATGCGATGCTC 1086
Qy 347 LeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAla 366
Db 1087 GTCCATTTTGGCATAGTTCGCAAGCTCGGAGGTTCTCTAGAAAAATGAAGCT 1146
Qy 367 GlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAsp 386
Db 1147 CGGAACAAGTCTCGAAGTATGCTCAGAGATGATATAGACGACAGTGAAGCTTAACTT 1206
Qy 387 IleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuSerThrArgThr 406
Db 1207 GTTGAAGCTTTTATTTGGATTTGAAAAGGCTCTGGAAGTACTGAATGCTGTGTAGACCG 1266
Qy 407 ThrGlyGlnProLeuValAspTrpAspCysPheLysThrLeuValAsnSerPheLys 426
Db 1267 GTGGATCGGCACCTTTGTCATGACTGGGCTCGCTGAAAACCATGTTAAGGACTTTTGG 1326
Qy 427 AsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIle 446
Db 1327 ACACATTTGTTGATCTTCTCAGTATGGATGAACACATGAGTCCCTTTGCAAAATATC 1386
Qy 447 CysAsnMetGlyAlaAspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
Db 1387 TGCAATGTAGGAATAAAGAAAGACGAATGGCTGAAGCCTCAGCAGCAAGCTTGT 1440

RESULT 9
LOCUS AY133531
DEFINITION Arabidopsis thaliana At4g32940/F26P21_60 mRNA, complete cds.
ACCESSION AY133531
VERSION AY133531.1 GI:22137031
KEYWORDS FLI_CDNA.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 1485)
Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Banh,J., Bowser,L.,
Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lee,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
and Ecker,J.R.
Arabidopsis ORF clones
Unpublished
2 (bases 1 to 1485)
Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Banh,J., Bowser,L.,
Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lee,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
and Ecker,J.R.
Direct Submission
Submitted (17-JUL-2002) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL CDNA (RAFL CDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.
The Salk, Stanford, PGECC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RAFL CDNA: Kim,C.J.,
Chen,H., Cheuk,R., Shinn,P., Banh,J., Bowser,L., Chang,E.,
Dale,J.M., Goldsmith,A.D., Jones,T., Kamiya,A., Karlin-Neumann,G., Lam,B.,
Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
Yamada,K., Yamamura,Y., Yu,G., Yu,S., Davis,R.W., Theologis,A.,
and Ecker,J.R.
Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.
FEATURES
source
1. 1485
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="4"
/clone="U14418"
/note="This clone is in pUNI 51
ecotype: Columbia"
1. 1485
/note="gamma-VPE (vacuolar processing enzyme)"
/codon_start=1
/product="At4g32940/F26P21_60"
/protein_id="AAM91361.1"
/db_xref="GI:22137032"
/translation="MATTMTRVSCVVLVLLSLVAVSAARSGPDDVILKPSASRF
PRAENDDDSNGTRWAVLVAGSSGYNRYHQADICLAYOLLRKGLKEENIVVFMVD
DIANNYENPRGCTIISPHGKDYGVKPKDYTDGDDVNDVNLFAVLGDKTKAVKGSCK
VYDGPNDHIFLEFYSDHCGPGLCMPTSPYLXANDVNLKKAALCYKGLYSVLA
CEGSIFFGLLEGLNIYATTAASNESSMGTCGPEPSPPEYETCLGDLXSVAM
EDSCHNLOTETLHQOVELKRRTPAPUCYSYGVSHVMOYCDVGISKDLUDLYNGTNPN
DNFTFADANSKPPSRVTRQADLYHFWEATRRAPGSARKTCAQKQVLEAMSHRLH
IDNSVLVGLKILGSRGPEVLNRYKVRGQPLVDVNDNCKNQVRAFERRHCGSLSQYGI
KIMRSFANICNAGIOMEAEASQACTTTLTPGWSSLNRGFSA"
BASE COUNT 409 a 308 c 365 g 403 t
ORIGIN
Alignment Scores:
Pred. No.: 3.65e-101 Length: 1485
Score: 1246.00 Matches: 233
Percent Similarity: 69.89% Conservative: 71
Best Local Similarity: 53.56% Mismatches: 121
Query Match: 50.77% Indels: 10
DB: 8 Gaps: 4

```

1150 CTTGAAGCCATGCTCACAGACTTCATATTTGACATAGCGTGATCTCGTCGAAAAATC 1209

391 SerValIysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro 410

1210 TTCTTTGGCATTTTCGAGAGGCTCTGAAGTGCTAAACAAAGTAGCGTCTGCTGGCAACCT 1269

411 LeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430

1270 CTAGTCGATGACTGGAACCTCCCTTAAAAATTCAGGTGAGGACTTTTCGAGAGGCACTGTGA 1329

431 AlaThrValHisThrGlyLeuLysThrThrGlyAlaLeuAlaAsnIleCysAsnMetGly 450

1330 TCCTGCTCAGTACCGGTATTCAGACACATGAGGCTCTTTGCAAACTCTGCAATGCAGG 1389

451 ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer 465

1390 ATTCAATGAGCAATGGAGGAGGAGCTTCACAGGCTTGTTACC 1434

RESULT 10
AY059104
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
The Salk, Stanford, PCEC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Narusaka, M., Nguyen, E., Palm, C. J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S. E., Shinozaki, K., Davis, R. W., Ecker, J. R. and Theologis, A.
Submitted (03-OCT-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

AY059104
1516 bp mRNA linear
PLN 18-APR-2002
Arabidopsis thaliana putative vacuolar processing enzyme gamma-VPE (At4g32340) mRNA, complete cds.
AY059104
GI:16323431
FLI cDNA.
Arabidopsis thaliana.
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1516)
Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Etgu, P., Lee, J. M., Toriumi, M., Yu, G., Brooks, S., Chao, Q., Chen, H., Karlin-Neumann, G., Kim, C., Lam, B., Miranda, M., Nguyen, M., Palm, C. J., Shinn, P., Southwick, A., Davis, R. W., Ecker, J. R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 1516)
Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Narusaka, M., Nguyen, E., Palm, C. J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S. E., Shinozaki, K., Davis, R. W., Ecker, J. R. and Theologis, A.
Direct Submission
Submitted (03-OCT-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PCEC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada, K., Banh, J., Banno, F., Chang, E., Dale, J. M., Goldsmith, A. D., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Wu, H. C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koese, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Narusaka, M., Nguyen, E., Palm, C. J., Shinn, P., Southwick, A., Tracy, S. E., Davis, R. W., Ecker, J. R. and Theologis, A.

Yamada, K. (SSP/PCEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PCEC) contributed equally to this work as PIs.
Location/Qualifiers
I. 1516
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

FEATURES
source

Qy	216	LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp	235
Db	685	CTTCTCGAGGTTTGAACATCTATGCCCAACATGCATCAACGCCGGAAGACAGTGG	744
Qy	236	GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly	255
Db	745	GGTACTATTCCCTGGAGAGAACCCAGTCCTCCACCGGAGTATGAACATTTGTTAGT	804
Qy	256	AspThrPheSerIleSerTyrLeuGluAspSerAspLeuHisAspMetSerLysGluThr	275
Db	805	GACTTGTACAGTGTGTGTGATGGAAGATAGTGTGATGCACAAATTTACAGACTGACACT	864
Qy	276	LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro	292
Db	865	CTGCACCAAGCAATATGAACTTGTGAAGAGGAGACTGCACCTGTTTGGGTACTCTTAT---	921
Qy	293	GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSer	312
Db	922	---GGTCTCATGTCATGCAATATGCGATGTAGGAATTAGCAAGGATTAATCTCGATCTT	978
Qy	313	TyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSer	330
Db	979	TATATGGGAACAAACCCCTGCCAATGACAATTTTACCTTTGCGGATGCGGAATTCATAAG	1038
Qy	331	SerProIleSerAsnSerGlyLeuValIasnProArgAspIleProLeuLeuTyrLeuGln	350
Db	1039	CCACCT-----TCAAGAGTTACAAACCCAGCTGATGCAGATCTTGTTCATTTTGG	1089
Qy	351	ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu	370
Db	1090	GAATAAGTACCCGAAAGACCAAGAGGTTCAGCAAGAAAAACAAGAGCTCAGAAACAAGTA	1149
Qy	371	LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu	390
Db	1150	CTTGAAGCCATGCTCTCACAGACTTCATATTTTGACAATAGCGTGATCTCGGAAAAATC	1209
Qy	391	SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro	410
Db	1210	TTGTTTGCAATTTCCGAGAGGTGCTGAAGTGCTAAACAAAGATACGCTGCTCGGGCAACT	1269
Qy	411	LeuValAspAspTyrAspCysPheLysThrLeuValIasnSerPheLysAsnHisCysGly	430
Db	1270	CTPATGCGATGACTCGGAATCGCTTTAAAAATCAGGTGAGAGCTTTCGAGAGCGACTGTGGA	1329
Qy	431	AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly	450
Db	1330	TGCGTGTCTCAGTACGGTATCAAGCACATCAGGCTCTTTTCAAACNTCTGCAATGCAGGG	1389
Qy	451	ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer	465
Db	1390	ATTCAAAATGGAGCAATGGAGGAGGAGCGCTTACAGGGCTTTGTACC	1434

RESULT 11	1760 bp	linear	PLN 08-OCT-2001
AF424619	Arabidopsis thaliana AT4G32940/F26P21_60	mrna	complete cds.
LOCUS	AF424619		
DEFINITION	AF424619		
ACCESSION	AF424619.1		
VERSION	GI:15983488		
KEYWORDS	FLI_CDNA.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.		
AUTHORS	Shinn, P., Chen, H., Cheuk, R., Kim, C. J., Koesea, E., Meyers, M. C., Banh, J., Bowser, L., Carninci, P., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jiang, P. X., Jones, T. M., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Liu, S. X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C. S., Palm, C. J., Pham, P. K., Quach, H. L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C. C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R. W., Theologis, A., and Eckstein, J. R.		

Qy 411 LeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430
Db 1342 CTAGTCGATGACGTCGAACTGCCTTAAATACAGGTCGAGGCTTCGAGAGGACGTCGTGA 1401
Qy 431 AlaThrValHisThrGlyLeuValAsnThrGlyAlaLeuValAsnIleCysAsnMetGly 450
Db 1402 TCCTGTCATGACGTCGATCAAGCAGCAGTCCTTTGCAACATCTGCAATGACAGG 1461
Qy 451 ValAspValLysGlnThrValSerAlaIleGluAlaCysSer 465
Db 1462 ATCAAAATGAGCAAAATGAGGAGGAGGCTTCACAGGCTGTGACC 1506

RESULT 12
AF370160
LOCUS
DEFINITION
Arabidopsis thaliana putative vacuolar processing enzyme gamma-VPE
(At4g32940) mRNA, complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT
TITLE
JOURNAL
COMMENT
FEATURES
source

1779 bp mRNA linear PIN 23-APR-2002
Arabidopsis thaliana putative vacuolar processing enzyme gamma-VPE
(At4g32940) mRNA, complete cds.
AF370160.1 GI:13877794
FLU CDNA.
Arabidopsis thaliana.
Arabidopsis thaliana.
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: eudicotyledons: Core eudicots:
Rosidae: eurosids II: Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1779)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Chou, R., Hayashizaki, I., Ishida, J., Lam, B.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R., and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2. (bases 1 to 1779)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G.,
Bowser, L., Carninci, P., Chen, H., Chou, R., Hayashizaki, I.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Kim, C., Koeseema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M.,
Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M.,
Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R., and
Theologis, A.
Direct Submission
Submitted (18-APR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X.,
Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D.,
Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G., Bowser, L.,
Chen, H., Chou, R., Jones, T., Karlin-Neumann, G., Kim, C., Koeseema, E.,
Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J.,
Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R., and Theologis, A.
Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
contributed equally to this work as PIs.
Location/Qualifiers
1. .1779
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="4"
/clone="RAFL05-09-J06 (R10153)"
/note="this clone is in a modified pBluescript vector
(Lambda ZAP) as a xhoI/sstI insert."

ecotype: Columbia"
1. .1779
/gene="At4g32940"
1. .74
/misc_difference 2
/gene="At4g32940"
/note="compared to genomic sequenc"
/replaces="c"
75. .1559
/gene="At4g32940"
/codon_start=1
/evidence-experimental
/product="putative vacuolar processing enzyme gamma-VPE"
/protein_id="AAK43975.1"
/db_xref="GI:13877795"
/translation="MATTMTVRVSVGVLLVLSVAVSAARSPPDDVIKLPQASRF
EPAENNDDNSGTSRWALVACSSGWNRYRHQADICHAYQLLRKGLKEENIVVFMVD
DIANNYENRPGTIINSHPGKDYQVCPKDTGDDVNDVNLFAVLGDKTAVKGSQK
VVDGPNDFHIFYSDHGGPVLGMPSTPYLYANDLNDVLLKHALGTYKSLVFLVLEA
CESGSIFFEGLLPEGLNIYATTAENAEESWMTYCPGEPSPEYETCLGLDYSVAVM
EDSGMNLQETLHQYELVKRTPAPVSGYSHVMOYGDVGIKADKDLVYMGNTNAN
DNFTFADANSLKPPSRVTNORDADLVHPKRYKAPGSAKTEAKQKDLVLEMSIRLH
INDSVLLGKILFGISRGPEVLNKRVSAGOPLVDDMNCLEKQVFAFERHCISLSOYGI
KWRSPANTCNAGIQMEQNEEAASQACTTLTPGTPSSLNKGFSA"
1560. .1779
3'UTR
misc_difference
/gene="At4g32940"
/gene="At4g32940"
/note="not present in genomic sequence"
BASE COUNT 509 a 360 c 407 g 503 t
ORIGIN
Alignment Scores:
Pred. No.: 4.64e-101 Length: 1779
Score: 1246.00 Matches: 233
Percent Similarity: 69.89% Conservative: 71
Best Local Similarity: 53.56% Mismatches: 121
Query Match: 10 Indels: 10
DB: 8 Gaps: 4
US-09-934-066-2 (1-466) x AF370160 (1-1779)
Qy 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
Db 219 GAAAGAGGAGCAGCATCTTAACCTCCGCTACTAGTGGCTGTCTTAGTCGCCGATCTAGC 278
Qy 56 GluTyThrAsnTrpArgHisGlnAlaAspIleCysHisAlaTyGlnIleLeuArgLys 75
Db 279 GGATATTGGAATTACAGGCATCAGCTCATATATCCATGCCCTATCAACTCTCTAGGAAA 338
Qy 76 GlyGlyLeuLysAspGluAsnIleValPheMetTyThrAspAspIleAlaPheSer 95
Db 339 GTGTGATTGAAGAGAGAGGAATATTGTGCTATTCATGATGATGATATTCCTAACCAATTAC 398
Qy 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyLysGly 115
Db 399 GAGATCCCAAGGCGCTGGAACCATTTATCAACAGCCCTCATGGAAAGATGTCTATCAAGGA 458
Qy 116 ValProLysAspTrpThrLysGluAlaValAsnValGlnAsnPheTyThrAsnValLeu 135
Db 459 GTTCCCAAGATTTATCATGAGATGATGTCATTAATGATTAATCTATTTGCTGTGATCCTT 518
Qy 136 GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAsp 155
Db 519 GGAGACAAAACGCTGCTTAAAGGGGAGAGTGGGAAGGTGTGGATAGTGGCTCTAATGAT 578
Qy 156 AsnIlePheIleTyThrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
Db 579 CATATCTTCATATTCACAGTGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
Qy 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195

collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN

.....

USA
RTKEN Genomic Sciences

center (GSC) members carried out the

Search completed: May 29, 2003, 19:19:32
Job time : 3147 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:08:47 ; Search time 378 Seconds
(without alignments)
9293.961 Million cell updates/sec

Title: US-09-934-066-1
Perfect score: 1560
Sequence: 1 ctcaagaatcagattca.....gaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1518.8	97.4	1541	21 AAC35910	Arabidopsis thalia
2	1401	89.8	1401	21 AAC50984	Arabidopsis thalia
3	454.8	29.2	1485	21 AAC42934	Arabidopsis thalia
4	452.8	28.0	1742	14 AAQ50570	Asparaginylenoep
5	451	28.9	1323	14 AAQ50560	Asparaginylenoep
6	442	28.3	1323	14 AAQ50562	Asparaginylenoep
7	442	28.3	1749	14 AAQ50574	Asparaginylenoep
8	441.6	28.3	1323	14 AAQ50561	Asparaginylenoep
9	441.6	28.3	1830	14 AAQ50572	Asparaginylenoep

10	345.6	22.2	1910	14 AAQ50579	Asparaginylenoep
11	345	22.1	1323	14 AAQ50559	Asparaginylenoep
12	343.4	22.0	1323	14 AAQ50563	Asparaginylenoep
13	323.6	20.7	1640	14 AAQ50575	Asparaginylenoep
14	319.2	20.5	1152	14 AAQ50564	Asparaginylenoep
15	305.4	19.6	894	14 AAQ50565	Asparaginylenoep
16	302.4	19.4	813	14 AAQ50566	Asparaginylenoep
17	281	18.0	753	14 AAQ50569	Asparaginylenoep
18	229.2	14.7	1453	14 AAQ50573	Open reading frame
19	198.8	12.7	1128	19 AAV29038	Human protein comp
20	198.8	12.7	1749	19 AAV29037	Osteoclast inhibit
21	198.8	12.7	1936	19 AAV44686	Human FHD02 cDNA
22	198.8	12.7	2030	22 AAC68662	Human adrenal gland
23	190.8	12.2	1855	18 AAT66381	Novel cysteine pro
24	190.8	12.2	1855	20 AAV69241	Human cysteine pro
25	190.8	12.2	1855	21 AAZ44095	Asparaginylenoep
26	183.4	11.8	1250	14 AAQ50571	DNA encoding novel
27	143	9.2	2373	23 AAS65199	DNA encoding novel
28	133.6	8.6	1336	24 ABK48384	Enhanced expressio
29	75.6	4.8	423	16 AAT35112	DNA sequence from
30	73.2	4.7	2135	22 AAC83228	Arabidopsis thalia
31	55.8	3.6	1490	21 AAC36486	Arabidopsis thalia
32	50	3.2	1491	21 AAC45045	Asparaginylenoep
33	49.2	3.2	329	14 AAQ50578	Asparaginylenoep
34	45	2.9	1492	19 AAV38747	cDNA nISP15-1492 e
35	45	2.9	1492	19 AAV38748	Complementary stra
36	45	2.9	5219	24 ABL49351	Human polynucleoti
37	44.6	2.9	1664976	19 AAV21209	Methanococcus jann
38	44.2	2.8	620	7 AAN60479	Encodes P. falcipa
39	43.6	2.8	15061	22 AAL04226	Human reproductive
40	43.6	2.8	15061	23 ABK42104	Genomic sequence #
41	43.4	2.8	1405	22 AAC83230	DNA sequence from
42	42	2.7	17137	24 ABL32191	Human immune syste
43	41.4	2.7	1524	21 AAA70107	Plasmodium falcipa
44	41.4	2.7	168575	22 AAH21613	Human hypocrerin r
45	41	2.6	294	22 AAH70082	Human cervical can

ALIGNMENTS

RESULT 1
AAC35910
ID AAC35910 standard; DNA; 1541 BP.
XX
AC AAC35910;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 11859.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0128845.

[illegible]

KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
QS
PW EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148545.
PR 13-AUG-1999; 99US-0148545.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149173.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR	10-SEP-1999;	99US-0153070.	QY	400	ACTAAGAAGCTGTTAATGTTTCAAACTTCTACAAATGTGTTTACTTGGAAATGAAGTGGC	459
PR	13-SEP-1999;	99US-0153758.	Db			
PR	15-SEP-1999;	99US-0154018.	Db	361	ACTAAGAAGCTGTTAATGTTTCAAACTTCTACAAATGTGTTTACTTGGAAATGAAGTGGC	420
PR	16-SEP-1999;	99US-0154039.				
PR	20-SEP-1999;	99US-0154779.	QY	460	GTCACAGAGAAATGCGAAAGTTCTGAAAAGTGGTCTTAATGATATATCTTCTCATCTAT	519
PR	22-SEP-1999;	99US-0155139.	Db			
PR	23-SEP-1999;	99US-0155486.	Db	421	GTCACAGAGAAATGCGAAAGTTCTGAAAAGTGGTCTTAATGATATATCTTCTCATCTAT	480
PR	24-SEP-1999;	99US-0155659.				
PR	28-SEP-1999;	99US-0156458.	QY	520	TATGCTGACCATGGAGCTCCTGGGCTTAAATACCGATGCCACTGGTGTGAAGTTATGGCA	579
PR	04-OCT-1999;	99US-0157117.	Db			
PR	05-OCT-1999;	99US-0157753.	Db	481	TATGCTGACCATGGAGCTCCTGGGCTTAAATACCGATGCCACTGGTGTGAAGTTATGGCA	540
PR	06-OCT-1999;	99US-0157865.				
PR	07-OCT-1999;	99US-0158029.	QY	580	AAAGATTTCAATGAAGTCTTGGAGAAGATGATGAAGAAAAAATACAAACAAGATGTG	639
PR	08-OCT-1999;	99US-0158232.	Db			
PR	12-OCT-1999;	99US-0158369.	Db	541	AAAGATTTCAATGAAGTCTTGGAGAAGATGATGAAGAAAAAATACAAACAAGATGTG	600
PR	13-OCT-1999;	99US-0159293.				
PR	13-OCT-1999;	99US-0159294.	QY	640	ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTC	699
PR	13-OCT-1999;	99US-0159295.	Db			
PR	14-OCT-1999;	99US-0159329.	Db	601	ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAATCTC	660
PR	14-OCT-1999;	99US-0159330.				
PR	14-OCT-1999;	99US-0159331.	QY	700	AACATATACGAGTACTGCTGCTAAATTTCTAAAGAGAGCAGCTGGGAGTTTACTGTCTT	759
PR	14-OCT-1999;	99US-0159637.	Db			
PR	14-OCT-1999;	99US-0159638.	Db	661	AACATATACGAGTACTGCTGCTAAATTTCTAAAGAGAGCAGCTGGGAGTTTACTGTCTT	720
PR	18-OCT-1999;	99US-0159584.				
PR	21-OCT-1999;	99US-0160741.	QY	760	GAGTCATATCT	819
PR	21-OCT-1999;	99US-0160767.	Db			
PR	21-OCT-1999;	99US-0160768.	Db	721	GAGTCATATCT	780
PR	21-OCT-1999;	99US-0160770.				
PR	21-OCT-1999;	99US-0160814.	QY	820	TCCTGGCTTGAGGACAGTACTGCTGCTAAATTTCTAAAGAGAGCAGCTGGGAGTTTACTGTCTT	879
PR	21-OCT-1999;	99US-0160815.	Db			
PR	22-OCT-1999;	99US-0160980.	Db	781	TCCTGGCTTGAGGACAGTACTGCTGCTAAATTTCTAAAGAGAGCAGCTGGGAGTTTACTGTCTT	840
PR	22-OCT-1999;	99US-0160981.				
PR	22-OCT-1999;	99US-0160989.	QY	880	CACGTTGTAAAGAGAGATAGGATCTGATGTACAGAGACTTCTCATGTATGCCGTTTC	939
PR	25-OCT-1999;	99US-0161404.	Db			
PR	25-OCT-1999;	99US-0161405.	Db	841	CACGTTGTAAAGAGAGATAGGATCTGATGTACAGAGACTTCTCATGTATGCCGTTTC	900
PR	25-OCT-1999;	99US-0161406.				
PR	26-OCT-1999;	99US-0161359.	QY	940	GGACAGAGAGATGCTTTAAAGATATCTTTCCCTCTTACATTTGGAAGAAATCCTGAAATC	999
PR	26-OCT-1999;	99US-0161360.	Db			
PR	26-OCT-1999;	99US-0161361.	Db	901	GGACAGAGAGATGCTTTAAAGATATCTTTCCCTCTTACATTTGGAAGAAATCCTGAAATC	960
PR	28-OCT-1999;	99US-0161920.				
PR	28-OCT-1999;	99US-0161992.	QY	1000	GATAACTTCACTTTTCACGGAATCCTTTTCCCTCACCAATCTCTAATCTCGGCTTGGTCAAT	1059
PR	28-OCT-1999;	99US-0161993.	Db			
PR	29-OCT-1999;	99US-0162142.	Db	961	GATAACTTCACTTTTCACGGAATCCTTTTCCCTCACCAATCTCTAATCTCGGCTTGGTCAAT	1020
			QY	1060	CCGCGCATATTCCTCTCTATACCTCCAGAGAAGATTTCAAAAAGCTTCCCAATGGGATCA	1119
			Db			
			Db	1021	CCGCGCATATTCCTCTCTATACCTCCAGAGAAGATTTCAAAAAGCTTCCCAATGGGATCA	1080
			QY	1120	CTTGAAGCAAGAGCTCAGAAATTTGCTTGACGAAAAGATCATAGGAACAATATC	1179
			Db			
			Db	1081	CTTGAAGCAAGAGCTCAGAAATTTGCTTGACGAAAAGATCATAGGAACAATATC	1140
			QY	1180	GATCAGAGCATTTACAGACATTTCTGGGCTTTTCAAGTAAACAAACCAATGTCTTAAATCTC	1239
			Db			
			Db	1141	GATCAGAGCATTTACAGACATTTCTGGGCTTTTCAAGTAAACAAACCAATGTCTTAAATCTC	1200
			QY	1240	TTAACTTCCACAAGAACACAGAGAGCTCTTCTTGACAGCATTTGGGATTTGCTTCAAGACT	1299
			Db			
			Db	1201	TTAACTTCCACAAGAACACAGAGAGCTCTTCTTGACAGCATTTGGGATTTGCTTCAAGACT	1260
			QY	1300	CTAGTTAATAGCTTCAAGAATCAGCTGCGGTGCAACGGTGCATTTAGGGATTTGAAGTATACA	1359
			Db			
			Db	1261	CTAGTTAATAGCTTCAAGAATCAGCTGCGGTGCAACGGTGCATTTAGGGATTTGAAGTATACA	1320
			QY	1360	GGAGCGCTTGCAATATCTGCAATATGGAGTGTGTAAGCAACCTGTTTCAGCCATT	1419
			Db			
			Db	1321	GGAGCGCTTGCAATATCTGCAATATGGAGTGTGTAAGCAACCTGTTTCAGCCATT	1380
			QY	1420	GAACAAGCTTGTTCGATGTAA	1440
			Db			
				1381	GAACAAGCTTGTTCGATGTAA	1401

RESULT 3
AAC42934
ID AAC42934 standard; DNA; 1485 BP.
XX AC AAC42934;
XX AC AAC42934;
DT 17-OCR-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37409.
XX DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137428.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146398.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR	20-AUG-1999;	99US-0149929.	QY	321	CTCGAGAACTCTAGGCGCTGGAGTTATCATATAAACCAGATGGAGAAAGATGTTTATAA	380
PR	23-AUG-1999;	99US-0149902.	Db			
PR	23-AUG-1999;	99US-0149930.	Db	321	TTACGAGAAATCCAAAGCCTGGAACCATATCAACAGCCCTCATGGAAGAAATGCTATCA	380
PR	26-AUG-1999;	99US-0150566.	QY			
PR	26-AUG-1999;	99US-0150884.	QY	381	AGGAGTTCCTAAGAGCTACACTAAAGAGCTGTAATGTTCAAACTTCTTACAAATGTTT	440
PR	27-AUG-1999;	99US-0151065.	Db			
PR	27-AUG-1999;	99US-0151066.	Db	381	AGGAGTTCCTAAGAGCTTATCTGAGATGATGCAATGTTGTAATCTATTTCTGTGAT	440
PR	27-AUG-1999;	99US-0151080.	QY			
PR	30-AUG-1999;	99US-0151303.	QY	441	ACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATGGCAAGTTTGTGAAAAATGGCTCTAA	500
PR	31-AUG-1999;	99US-0151438.	Db			
PR	01-SEP-1999;	99US-0151930.	QY	441	CTTTGGAGACAAACTGCTGTTAAAGGGGAAGTGGGAAGTTTGGATAGTGGCTCTAA	500
PR	07-SEP-1999;	99US-0152363.	QY	501	TGATAATATCTTCATCTATTTATCTGACCATGAGAGCTCTGGCTTAAATAGCGATGCCAC	560
PR	10-SEP-1999;	99US-0153070.	Db			
PR	13-SEP-1999;	99US-0153758.	Db	501	TGATCATATCTTCATATTTACAGTGACCATGTTGGTCTCTGGAGTTCTTTGGGATGCCAC	560
PR	15-SEP-1999;	99US-0154018.	QY			
PR	20-SEP-1999;	99US-0154039.	QY	561	TGCTGATGAAGTTATGCAAAAGATTTCAATGAAGTCTTGGAGAAGATGATAAGAGAAA	620
PR	22-SEP-1999;	99US-0155139.	Db			
PR	23-SEP-1999;	99US-0155486.	Db	561	TTCTCTTACCTATATGCAATGATCTCAATGATGTTTGAAGAAGAAACATGCTTTAGG	620
PR	24-SEP-1999;	99US-0155659.	QY			
PR	28-SEP-1999;	99US-0156458.	QY	621	AAATATACAAAGATGGTATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGG	680
PR	29-SEP-1999;	99US-0156596.	Db			
PR	04-OCT-1999;	99US-0157117.	Db	621	AACATATAAAGCTTGGTGTGTTTATCTCGAAGCTTGGGAATCTGGAAGTATCTTTGAAGG	680
PR	05-OCT-1999;	99US-0157753.	QY			
PR	06-OCT-1999;	99US-0157865.	QY	681	GATTTTAAAGAAAATCTCAACATATACGCAGTACTGCTGCTTAATTTCTAAAGAGACAG	740
PR	07-OCT-1999;	99US-0158029.	Db			
PR	08-OCT-1999;	99US-0158232.	Db	681	GCTTCTCTCGAGGGTTTGAACATCTATGCCACAACATGTCATAAACCGCCGAAGAACAG	740
PR	12-OCT-1999;	99US-0158369.	QY			
PR	13-OCT-1999;	99US-0159293.	QY	741	CTGGGAGTTTACTGTCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	800
PR	13-OCT-1999;	99US-0159294.	Db			
PR	13-OCT-1999;	99US-0159295.	Db	741	TTGGGGTACCTATTTGGCCCTGGAGAGAACCCAGTCTCCACCGGAGTATGAAATCTGTTT	800
PR	14-OCT-1999;	99US-0159329.	QY			
PR	14-OCT-1999;	99US-0159330.	QY	801	CGGGATACATTTAGCATCTCTTTGGGCTTGAGGACAGTACCTTCATGACATGAGCAAGA	860
PR	14-OCT-1999;	99US-0159331.	Db			
PR	14-OCT-1999;	99US-0159637.	Db	801	AGTGACTTGTACAGTGTCTTGGATGGAAGATAGTGTATGCACAATTTACAGACTGA	860
PR	18-OCT-1999;	99US-0159384.	QY			
PR	21-OCT-1999;	99US-0160741.	QY	861	GACTTTGGAGCAACAATACCACGTTGTAAGAGAAGAGTAGGATCTGATGTAC---CAGA	917
PR	21-OCT-1999;	99US-0160767.	Db			
PR	21-OCT-1999;	99US-0160768.	Db	861	GACTCTGCACCAATATGAACCTTGTGAAAGAGGAGCTGCACCTTGGTACTCTTA	920
PR	21-OCT-1999;	99US-0160770.	QY			
PR	21-OCT-1999;	99US-0160814.	QY	918	GACTTCTCATGTATGCGGTTTCGGAACAGAGAGTCTTAAAGATTTCTTCTCTCTTA	977
PR	21-OCT-1999;	99US-0160815.	Db			
PR	22-OCT-1999;	99US-0160980.	Db	921	TGTTTCTCATGTATGCAATATGGCATGTAGGAATTTAGCAAGATTAATCTCGATCTTTA	980
PR	22-OCT-1999;	99US-0160981.	QY			
PR	22-OCT-1999;	99US-0160989.	QY	978	CATTGGAAGAAATCCTGAAACGATACCTTTCACCTTTCACGGAATCTTTTCTCCACCAAT	1037
PR	25-OCT-1999;	99US-0161404.	Db			
PR	25-OCT-1999;	99US-0161405.	Db	981	TATGGGAACAACCCCTGCCAATTTTACCTTTTGGCGGATGCGGAAT--TCACATAA	1037
PR	25-OCT-1999;	99US-0161406.	QY			
PR	26-OCT-1999;	99US-0161359.	QY	1038	CTCTAATTTCTGGCTTGGTCAATCCGGCGGATATTTCTCTCTCTATACCTCCAGAGAAAT	1097
PR	26-OCT-1999;	99US-0161360.	Db			
PR	26-OCT-1999;	99US-0161361.	Db	1038	GCCACCTTCAAGAGTTTACAACCCAGCGTGATGCGAGATCTTTGTTTCATTTTGGGAAAAGTA	1097
PR	28-OCT-1999;	99US-0161920.	QY			
PR	28-OCT-1999;	99US-0161920.	QY	1098	TCAAAAAGCTCCAATGGGATCACTTGAAGCAAGAGAGCTCAGAAGAAATTTGCTTGACGA	1157
PR	28-OCT-1999;	99US-0161992.	Db			
PR	28-OCT-1999;	99US-0161993.	Db	1098	CCGAAAGCACCAGAGGTTTCAAGAGAAACAGAGCTCAGAAGCAAGTACTTTGANGC	1157
PR	29-OCT-1999;	99US-0162142.	QY			
PR	29-OCT-1999;	99US-0162142.	QY	1158	AAAGAATCATAGGAACAAATCGATCAGAGCATTTACAGACATTTCTGCGGCTTTTCAGTTAA	1217
PR	29-OCT-1999;	99US-0162142.	Db			
PR	29-OCT-1999;	99US-0162142.	Db	1158	CATGCTCACAGACTTCATATTTGACAATAGCGTGATACTCTCGGAAATTTCTGTTGG	1217
QY	141	TGTTGAATCTAGGCAAGAGTGC	QY	1218	ACAAACCAATGCTTAAATCTTTAACTTCCCAAGAACAAACAGGACAGGCTCTTTGAGA	1277
Db			Db			
QY	141	TGCTGAAAGCAGGAGGATCTTAACTCCGCTAGGTGGCTGTTCTAGTCGCGGATC	Db	1218	CATTTCGAGAGGCTCCGAGTGTCTTAAACAAAGTACGCTCTGCTGGGCAACCTAGTCGA	1277
QY	201	AAATGAATATTAATACAGGATCAGGCTGACATATGCCAGCGTATCAGATCTCCG	QY	1278	CGATTGGGATTTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTGCGGCTGCAACGGT	1337
Db			Db			
QY	201	TAGCGATATTTGAATTTACAGGATCAGGCTGATATATGCCCTTCACTTCTGAG	QY	1337	TGACTGGAATGCTTCAAAATCAGGTGAGAGCTTTTCGAGAGGACACTGTGGATCGCTGTC	1337
QY	261	AAAAGCGGTTTAAAAGATGAAACATCATTTGTTTATGATGATGATATCGCGTTTC	QY	1338	GCATTACGAGTTGAAGTATACAGGAGCGCTTGGCAATATCTGCAATATGGAGTGGATGT	1397
Db			Db			
Db	261	GAAGGTGGATTGAAGAGAGAGATATTTGGTATTTCATGATGATGATGATGCTCAACA	Db	1338	TCAGTACGGTATCAAGACATAGGCTCTTTTGGCAACATCTGCAATCGAGGATTCAT	1397
			QY	1398	GAAGCAAACTGTTTTCAGCCCATTTGAACAAAGCTTGTTCGA	1435

Query Match 29.2% Score 454.8; DB 21; Length 1485;
 Best Local Similarity 60.5% Pred. No. 4.5e-107;
 Matches 785; Conservative 0; Mismatches 507; Indels 6; Gaps 2;

```

Db      1398 GGAGCAATGGAGGAGGAGCTTCACAGGCTGTGTACCA 1435
      | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 4
AAQ50570
ID: AAQ50570 standard; cDNA to mRNA; 1742 BP.
XX
AC AAQ50570;
XX
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase clone 101.
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
XX Key Location/Qualifiers
FH CDS 207..1529
FT /*tag= a
XX
XX JP05276960-A.
XX
XX 26-OCT-1993.
XX
XX 07-AUG-1992; 92JP-0231602.
XX
XX 07-FEB-1992; 92JP-0056023.
XX
XX (SYUZ/) SYUZO T.
XX
XX WPI: 1993-373587/47.
XX
XX P-PSDB; AAR43035.
XX
XX New gene for encoding asparaginyl endo-peptidase - comprises 8
XX specified DNA sequences
XX
XX Disclosure; Page 17-19; 35pp; Japanese.
XX
XX A gene encoding asparaginylendopeptidase is claimed.
XX 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
XX derived from a seed of Canavalia ensiformis which selectively
XX hydrolyses C-terminus amide bond of L-asparagine residue
XX (see AAR43033 and AAR43041).
XX The enzyme is useful for protein fragmentation and enzymatic
XX peptide synthesis.
XX The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
XX in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
XX by PCR.
XX
XX Sequence 1742 BP; 507 A; 338 C; 395 G; 502 T; 0 other;
XX
XX Query Match 29.0%; Score 452.8; DB 14; Length 1742;
XX Best Local Similarity 60.6%; Pred. No. 1.6e-106;
XX Matches 780; Conservative 0; Mismatches 502; Indels 6; Gaps 2;
XX
QY 146 AATCTAGGCACAGAGTGCAGGAGGACACAGGATGGCTGTTTGTCTGCTGATCAATG 205
   || | | | | | | | | | | | | | | | | | | | | | | | |
DB 190 AACCCACCCAGGAAACAGCAGGAGGACACAGGATGGCGCTTTTAACTGCTGGTCCCATG 249
XX
QY 206 AATATTATACAGCGATCAGCTGACATATGCCACGCGTATCAGATCTCCGAAAG 265
   || | | | | | | | | | | | | | | | | | | | | | | | |
DB 250 GTTACTGGAAATACAGGCAATCAGTCTGATGATGCCATCGCTCAAGTCTGAGGAAAG 309
XX
QY 266 GCGGTTTAAAGCAATCATCTGTTTATGATGATATGATATGATATGATATGATATGAT 325
   || | | | | | | | | | | | | | | | | | | | | | | | |
DB 310 GTGGCTGAAGAGAAATATGTTGTTTATGATGATGATGATGATGATGATGATGATGAT 369
XX
QY 326 AGAATCCTAGGCGCTGAGTTATCATTAATAAACACAGATGGAGAGATGTTTATAAGGAG 385
   || | | | | | | | | | | | | | | | | | | | | | | | |

```


XX DT 24-MAY-1994 (first entry)
 XX DE Asparaginylendopeptidase sequence (4).
 XX KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
 XX KW L-asparagine; primer; PCR; protein fragmentation;
 XX KW peptide synthesis; ss.
 XX OS Canavalia ensiformis.
 XX PA Canavalia ensiformis.
 XX PN JP05276960-A.
 XX PD 26-OCT-1993.
 XX PF 07-AUG-1992; 92JP-0231602.
 XX PR 07-FEB-1992; 92JP-0056023.
 XX PS (SYUZ/) SYUZO T.
 XX WIPI; 1993-373587/47.
 XX DR New gene for encoding asparaginyl endo-peptidase - comprises 8
 XX PT specified DNA sequences
 XX PS Claim 1; Page 13-14; 35pp; Japanese.
 XX CC A gene encoding asparaginylendopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ05057-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ050567-68, AAQ050576-77 and AAQ050583-90 were used
 CC in the isolation of the fragments given in AAQ050569-75 and AAQ050578-79,
 CC by PCR.
 XX SQ Sequence 1323 BP; 375 A; 255 C; 328 G; 365 T; 0 other;

Query Match 28.3%; Score 442; DB 14; Length 1323;
 Best Local Similarity 60.4%; Pred. No. 8.7e-104;
 Matches 765; Conservative 0; Mismatches 495; Indels 6; Gaps 2;
 169 GGCACAGATGGGCTGTTTGTGTCGATCAATGAATATATTAACACAGGCATCAG 228
 7 GGTACACAGTGGGCGATCTCTCGCGGTTCGAATGGCTACTGGAATACAGGCATCAG 66
 229 GCTGACATATGCCACGCTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAACATC 288
 67 GCTGATATTTGTCATGCGTATCAACACTGAGGAAAGGTGGCGTGAAGAAGAAATATT 126
 289 ATTGCTTTTATGATGATATCGCGTTTCTCGGAGAACTCTAGGCTTGAGTTATC 348
 127 ATTTCTTTTATGATGATGATGATGCTTCAATGAGGAGAACCAAGACCTGGGTCATC 186
 349 ATTAATAACACAGATGAGAGATGTTTATAAGAGGTTCTTAAGACTACATCAAGAA 408
 187 ATTAACAACACAGATGGAGCTGATTTATGAGGAGTTCCAAGGATTTACACTGGCGAA 246
 409 GCTGTTAATGTTCAAACTTCTACATGTTTACTTGTGAAATGAAGTGGCGTCACAGA 468
 247 GATGTTACTGCTGAAACTTCTACGCTGTTTACTTGGAAATGAAGTGGCACTTAGCGGT 306
 469 GGAATGGCAAGTTTGGAAAGTGGTCTTAATGAATATCTTCACTATTATTGCTGAC 528
 307 GGCAGTGGGAAGTTTGGAAAGTGGTGGCGGATGATCGGATATTTGATTTATATGATGAT 366
 529 CATGAGCTCTGGCTTTAATAGCGATGGCCACTGGGTGATGAAGTTATGGCAAAAGATTTC 588
 367 CATGGAGTCCAGGAGTCTCGGATGGCTGGCGGCTTACTTATATGATCTGATCTG 426

QY 589 AATGAAGTCTTTGAGAGATGCATTAAGAGAAATAAATAACAAGATGGTGATCTATGTT 648
 DB 427 ATGATGCTTTGAAGAAAAAACATGCTTCGGAACATATAAACCCCTAGTATTTATCTT 486
 QY 649 GAAGCATGTGAATCAGGAAGTATCTTTGAAGGATTTTAAAGAAATCTCAACATATAC 708
 DB 487 GAGGATGTGAATCTGGAGTATCTTTGAAGGCTTCTTCTCCTGAAGATCTCAATATTTAT 546
 QY 709 GCAGTGAAGTCTGCTAAATTTCTAAAGAGAGAGCTGGGAGTCTTACTGTCTCTGAGTCATAT 768
 DB 547 GCAACAACAGCTTCGAATGCAAGAAAGAGAGCTGGGAAACATATTTGCCCGAGGAGTAT 606
 QY 769 CTTCTCTCTCTCTGAGATTTGGAATCTTCTCGGCATACATTTAGCATCTCTTGGCTT 828
 DB 607 CCCAGTCTCTCCCGAGAAATATTCAACTTGTCTGGAGAGCTGTACAGTGTGCTTGGATG 666
 QY 829 GAGGACAGTGAAGTCTTCAATGACATGAGCAAGAGAGCTTTGGAGCAACAATAACCA 885
 DB 667 GAAGACAGTGAAGAGCAACAATTTGCGAACTGAAACTTTGAACCAACAATAATAAATTTGGTT 736
 QY 886 GTAAAGAGAGAGTAGGATCTGATGTACAGAGAGCTTCTCATGTATGCTGCTTTCGGAACA 945
 DB 727 AAGAGAGAGGACCATTTAGTGGAGTTTCATCTATGCTCTCACGTGATGAGTATGTTGAT 786
 QY 946 GAGAAGATGCTTAAAGATTTATCTTCTCTTACATGGAAGAAATCTCTGAAACGATAC 1005
 DB 787 ATAGGCTCAGCGATGATGTTCTCTCTCTATATTTGGGTACAATCTCTGCTATGATAAT 846
 QY 1006 TTCACCTTTCAGCGAATCTTTTCTCTCAGCAATCTCTAATTTCTGCTTGGTTCATTCGCGC 1065
 DB 847 TTTACCTTTGTCGATGAAGAACTCTCTTG---AGGTCACTTTCAAAAGCAGTCAACCAACGG 903
 QY 1066 GATATCTCTGCTGATATACCTCCAGAGAAAGATTTCAAAAGCTCCAATGGGATCAGTTGAA 1125
 DB 904 GATGCTGATCTCATCTCCATTTCTGGGAGAAAGTTCGCAAGAGCTCTGAGGGTCTTCCCGAG 963
 QY 1126 AGCAAGAAGCTCAGAGAAATTTGCTTTCAGCAAGAAAGATCATAGGAACAAATTCGATCAG 1185
 DB 964 AAAATGCTGCTCAGAAACAAGTTGTGGAAGTAAATGTCTCACAGGATGCATATAGACAAC 1023
 QY 1186 AGCATATACAGACATTTCTCGGCTTTTCAGTTAAACAACCAATCTCTTAAATCTCTTAACT 1245
 DB 1024 GGTGTGGAACCTTATTTGGGAAGCTTTTATTTGGCATTTGAAAGGGTCCAAAGACTCTGGAT 1083
 QY 1246 TCCACAAGAAACAACAGCAGCCTCTTTGTAGAGGATTTGGGATTTGCTTCAAGACTCTAGTT 1305
 DB 1084 GCTGTAGACCGGCTGGAATGGCACTTTGATGACTGGGACTGCTGAAACCACTGTTA 1143
 QY 1306 AATAGCTTCAAGAAATCACTGCGGTGCAACGCTGCAATTTACGATTTGAAGTATACAGAGCG 1365
 DB 1144 AGGACATTTGAGACATATTTGTTGGATCTCTCTCAGTATGGATGAACATATGAGGTCC 1203
 QY 1366 CTTGCCAATATCTCAATATGGAGTGGATGTGAGCAAACTCTTTCAGCCTTGAACAA 1425
 DB 1204 TTTCCAAACATCTCCACGAGGAAATTAAGAATGACCAAAATGGCCGATGCTTCAGCAACA 1263
 QY 1426 GCTTGT 1431
 DB 1264 GCTTGT 1269
 RESULT 7
 AAQ050574
 ID AAQ050574 standard: cDNA to mRNA; 1749 BP.
 XX AC AAQ050574;
 XX XX
 XX 24-MAY-1994 (first entry)
 XX Asparaginylendopeptidase clone 107.
 XX DE Asparaginylendopeptidase; Canavalia ensiformis; seed;
 XX KW L-asparagine; primer; PCR; protein fragmentation;

specified DNA sequences

Claim 1; Page 12; 35pp; Japanese.

A gene encoding asparaginylendopeptidase is claimed.
8 DNA sequences are given (AAQ50559-66). The enzyme is a protease derived from a seed of Canavalia ensiformis which selectively hydrolyses C-terminus amide bond of L-asparagine residue (see AAR43033 and AAR43041).

The enzyme is useful for protein fragmentation and enzymatic peptide synthesis.

The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79, by PCR.

Sequence 1323 BP; 385 A; 246 C; 346 G; 346 T; 0 other;

Query Match 22.1%; Score 345; DB 14; Length 1323;

Best Local Similarity 56.3%; Pred. No. 8.6e-79;

Matches 692; Conservative 0; Mismatches 525; Indels 12; Gaps 2;

169 GGCACACGATGGCGTCTTGTAGTCTGGATCAAAATGAATATTATAACTACAGGCATCAG 228

7 GGAACCCGGTGGCGGCTCTCGTCTGCTTCAACCGGTTATGGAATATATAGGCATCA 66

229 GCTGACATATGCCACCGGTATCAGATATCTCCGAAAAGCGGTTTAAAAGATGAAAACATC 288

67 GCGGATGTGTCCATGCTTACCAGTGTCTGATATAAAGGTGGAGTCAAAAGAAAGAACAT 126

289 ATTGTGTTTATGATGATCATATCCGTTTCTCCGGAGATCTTAGCCCTGGAGTTATC 348

127 GTGGTGTGTTATGATGATGATATGCTTATAACGCCATGAATCCAGACCCGGGATCATC 186

349 ATTAATAAACACGATGGAGAGATGTTTATAAAGGAGTCTCTAAGGACTACACTAAAGAA 408

187 ATCAACCATCTCCAGGGCCAGACGTGTATGCTGTGTACCTAAGATTAACCCGGTCAG 246

409 GCTGTTATGTTCAAACTTCTACAAATGTGTACTTGGAAATGAAAGTGGCGTCAACAGGA 468

247 GACGTAACACCTGAGAAGCTATATCTGTCTATCTTGGGGACAAAGATAAGTAAAGT 306

469 GAAATGCAAGTGTGAAAGTGGTCTTAATGATATATCTTATCTATATATCTATCTAC 528

307 GGAAGTGGCAAGGTGATCAACAGCAATCCGGAGGATAGGATATTATATCTTACTCCGAT 366

529 CATGAGCTCTCGGCTTAATACGATGCCACTGTGTATGAAGTTATGGCAAAAGATTTC 588

367 CATGAGGTCCGGAGTCTTGGGATGCCAAGCACCATTCTGTTATGCCATGGATT 426

589 AATGAAGTCTTGGAGAAGATGCATGAAGAGAAAAAATACAAAGATGGTGTATCTATGTT 648

427 ATTGATGTTTGAAGAAGAAACATGCAAGTGGAGGCTACAAGGAGATGTTATATACATA 486

649 GAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAATCTCAACATATAC 708

487 GAAGCTTGTGAGGTGGGAGCATATTTGAGGGTATTAATGCCAAGGATCTGAATATTTAT 546

709 GCAGTGACTGTCTGCTTAATCTTAAGAGACAGCTGGGGAGTTTACTGCTCGAGTCATAT 768

547 GTGCAACTGCGTCAATGCCAAGACAGACATTTTGGAACTTATGTTCTGGGGATGAAT 606

769 CCTCTCTCTCTTCTGATTTGGAATGGAATCTGCTCGGCGATACATTTAGCATCTCTTTGGGTT 828

607 CCTCTCCACCAAGAGATGACGTAACCTTGCCTGGGGATTTTATACAGCGTTTCTTGGATG 666

829 GAGGACAGTACCTTCATGACATGAGCAAGAGACTTTGGAGCAACATACCACTTGTGA 888

667 GAAGATGTGAGACTCAACAATTAAGAGGGGAAACGGTACACAGCAATACCACTGCGGTA 726

889 AAGAGAAGAGTAGGATCTGATGTACCAAGAGAC-----TTCTCATGTATGCGGTTTCGGA 942

727 AGGAACCGACTTCAATCTTAACAGCTATAGGTTTGGTTCTCATGTATGCAATACGGT 786

QY 943 ACAGAGAAGATGCTTTAAAGATATATCTTCTCTTACATTTGGAAGAAATCTCTGAACGAT 1002
DB 787 GACACTAACTACTACTGCTGAAAGCTTTTACTTGTACCATGTTTGTATCTCTGCCACCGTG 846
QY 1003 AACCTTCACTTTACCGGAATCCTTTTCTTCCATCACAATCTCTAATTTCTGGCTTGGTCAATCG 1062
DB 847 AACCTTCTCTCC-----ACACAACGGCAACCTAGAAAGCTAAATGGAAGTTGTTAAACCAG 900
QY 1063 CGCGATATTTCTCTCTGCTATACCTCCAGAGAAGATTTCAAAAAGCTCCAAATGGGATCATT 1122
DB 901 AGAGATGCGAGGCTTTTCTCATGTGGCAATGTATCAGAGATCAAAACCATCAACCGNA 960
QY 1123 GAAACCAAGAGCTCAGAGAATTTGCTTGACGAAAAGAAATCATAGGAAACAAATCGAT 1182
DB 961 AAGAAGACTCATCTCTGGAACAGATTACAGAGACAGTGAAGCATAGGAATCATTTGGAT 1020
QY 1183 CAGAGCATTTACAGACATTTCTCGGCTTTTCAGTTAAACAACCAATGTCTTAATCTCTTA 1242
DB 1021 GCAGTGTGGAATTTGATTTGGAGTTTGTGTATGACCAAGAAAAGTTCTTCGGTTCTTA 1080
QY 1243 ACTTCCACAAGAACACAGGACAGCCTCTTGTAGACGATTTGGGATTTGCTTCAAGACTCTTA 1302
DB 1081 CATTCGCTGAGGCTCTCTGGTCTGCCCTAGTTGATGATTTGGACATGCTTTGAAATCTATG 1140
QY 1303 GTTAATAGCTTCAAGAATCACTGCGGTGCAACGGTGCATTTACGGATTTGAAGTATACAGGA 1362
DB 1141 GTTAGAGTGTTCGAAACTCACTGTGGTCACTGACTAGTATGGCATGGAACACATGCGG 1200
QY 1363 GCGCTTGCAATATCTGCAATATGGAGT 1391
DB 1201 GCATTCGCAAGCTTTGCAACAGCGCGT 1229

RESULT 12

AAQ50563

ID AAQ50563 standard; cDNA to mRNA; 1323-BP.

XX AAQ50563;

XX 24-MAY-1994 (first entry)

XX Asparaginylendopeptidase sequence (5).

XX Asparaginylendopeptidase; Canavalia ensiformis; seed;

KW L-asparagine; primer; PCR; protein fragmentation;

KW peptide synthesis; ss.

OS Canavalia ensiformis.

XX JP05276960-A.

XX 26-OCT-1993.

XX 07-AUG-1992; 92JP-0231602.

XX 07-FEB-1992; 92JP-0056023.

XX (SYUZ/) SYUZO T.

XX WPI; 1993-373587/47.

XX New gene for encoding asparaginyl endo-peptidase - comprises 8 specified DNA sequences

PS Claim 1; Page 14; 35pp; Japanese.

XX A gene encoding asparaginylendopeptidase is claimed.

CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively

CC hydrolyses C-terminus amide bond of L-asparagine residue

CC (see AAR43033 and AAR43041).

CC The enzyme is useful for protein fragmentation and enzymatic

CC peptide synthesis.

CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.

XX Sequence 1323 BP: 383 A; 247 C; 346 G; 347 T; 0 other;

Query Match 22.0%; Score 343.4; DB 14; Length 1323;

Best Local Similarity 56.2%; Pred. No. 2.2e-78;

Matches 691; Conservative 0; Mismatches 526; Indels 12; Gaps 2;

```
QY 169 GGCACACATGGCGTGTAGTTGCTGGATCAATTAATTAATTAACACAGGATCAG 228
    || || || || || || || || || || || || || || || || || || || || ||
Db 7 GGTACCCGGTGGCGGTTCTCGTGGCTTCAACCGTTATGGAATATAGGCATCAA 56
QY 229 GCTGACATATGCCAGCGGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAACATC 288
    || || || || || || || || || || || || || || || || || || || || ||
Db 67 GCGATGTGCCATGCTTACCATGCTGATGAAGGGTGGAGTCAAGAAAGAAACAT 126
QY 289 ATTGTGTTTATGATGATATGCGCTTTTCTCGGAGAACTCCTAGGCGTGGAGTTATC 348
    || || || || || || || || || || || || || || || || || || || || ||
Db 127 GTGGTGTATGATGATGATATGCTTATAGCCATGAATCCAGCCAGCCGAGTTCATC 186
QY 349 ATTAATAACACAGATGGAGAGATGTTTATAAGGAGTTCTTAAGGACTACACTAAAGAA 408
    || || || || || || || || || || || || || || || || || || || || ||
Db 187 ATCAACCATCTCAGGGGCCAGAGCTGTATGCTGGTACCTAAGGATTACACCGGTGAG 246
QY 409 GCTGTAAATGTTCAAACTTCTACAATGTCTTACTTGGAAATGAAAGTGGCGTCACAGGA 468
    || || || || || || || || || || || || || || || || || || || || ||
Db 247 GACGTAACACCTGAGAACCTATATGCTGCTCATCTTTGGGGACAAAGTAAAGTTAAAGT 306
QY 469 GGAATGCAAAAGTTGTGAAAGTGGTCTTAATGATAATATCTCATCTATTATGCTGAC 528
    || || || || || || || || || || || || || || || || || || || || ||
Db 307 GGAAGTGCAGAGGTGATCAACAGCAATCCGAGAGATAGGATATTTATATCTACTCCGAT 366
QY 529 CATGAGCTCTGCTGCTTAATAGCGATGCCACCTGGTGTGATGAAGTTATGCAAAAGATTC 588
    || || || || || || || || || || || || || || || || || || || || ||
Db 367 CATGAGGTCGCGAGTCTTGGATGCCAACGCCACCATCTCGTTACGCCATGGATTTT 426
QY 589 AATGAGTCTTGGAGAGATGCTACAGAGAAAATACAAAGATGGTGTATGTT 548
    || || || || || || || || || || || || || || || || || || || || ||
Db 427 ATTGATGTTTGAACAAGAACATGCAAGTGGAGGCTACAGGAGATGGTTATATACATA 486
QY 649 GAAGCATGTGAATCAGAGATGTTTGAAGGGATTTTAAAGAAAATCTCAACATATAC 708
    || || || || || || || || || || || || || || || || || || || || ||
Db 487 GAAGCTGTGAGTGGAGGATATTTGAGGGTATATGCCCCAAGGATCTGAAATATTTAT 546
QY 709 CAGTGTGCTGCTAATCTAAGAGAGAGCTGGGGATTTTACTGTCTCTGAGTCATAT 768
    || || || || || || || || || || || || || || || || || || || || ||
Db 547 GTGAACATCGCTCAAAATGCACAGAGAACAGTTTTGGAACTTATTTGCTCTGGGATGAAT 606
QY 769 CTTCTCTCTCTGAGATGGAATGTCTCGGCGATACATTTAGCATCTCTTGGCTT 828
    || || || || || || || || || || || || || || || || || || || || ||
Db 607 CTTCTCTCACCAGAGAGTACGTAACCTTGCTGGGGATTTATACAGCGTTTCTTGGATG 666
QY 829 GAGGACAGTGAACCTTCATGACATGAGCAAGAGACTTTGGAGCAACAATACCACTGTA 888
    || || || || || || || || || || || || || || || || || || || || ||
Db 667 GAAGATAGTGAGACTCAATCTAATAAGGGAACGCTACACAGCAATACCACTCGGTA 726
QY 889 AAGAGAGAGTATGATCTGATGATACAGAGAC-----TCTCATCTATGCCGTTCCGA 942
    || || || || || || || || || || || || || || || || || || || || ||
Db 727 AGGAACCGGACTTCAATTTCTAACAGCTATAGTGTGGTTCATCTGATGAATACGTT 786
QY 943 ACAGAGAGATGCTTAAAGATTTATCTTCTCTTACATGGGAAGAATCCTGAAACGAT 1002
    || || || || || || || || || || || || || || || || || || || || ||
Db 787 GACACTATCACTACTGCTGAAGAGCTTTACTTGTACCATGGTTTTGATCCTGCCACCGTG 846
QY 1003 AACCTTCACTTTTCAGGAATCCTTTTCTCACCACATCTCTAATTTCTGGCTTGGTCAATCCG 1062
    || || || || || || || || || || || || || || || || || || || || ||
Db 847 AACCTTCTCTC-----ACACAACGGCAACCTAGAAGCTTAAAGTGAAGATTTGTTAACCCAG 900
QY 1063 CGGATATTTCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCACTTT 1122
    || || || || || || || || || || || || || || || || || || || || ||
Db 901 AGAGATCGAGAGCTTTTGTGTTCTATGTCGCAAAATGTATCAGAGATCAAAACCATCAACCGGAA 960
```

```
QY 1123 GAAAGCAAAAGAGCTCAGAAAGAAATGCTTTGACGAAAAAGAAATCATAGGAAACAAATCCGAT 1182
    || || || || || || || || || || || || || || || || || || || || ||
Db 961 AAGAAGACTCACATCTCTGGAACAGATTTACAGACAGAGTGAAGCATAGGAATCATTTGGAT 1020
QY 1183 CAGAGCATTAACAGACATTTCTCGGCTTTTCAAGTTTAAACAACCAACATCTTTAAATCTCTTA 1242
    || || || || || || || || || || || || || || || || || || || || ||
Db 1021 GGCAGTCTGGAATTTGATTGGAGTTTGTCTATGACAGGAAAGATTTCTTCGGTTCTTA 1080
QY 1243 ACTTCCACAAGACAACAGGACAGACGCTCTTTGAGAGAGATTTGGATTTGCTTCAAGACTCTTA 1302
    || || || || || || || || || || || || || || || || || || || || ||
Db 1081 CATTCCTGGAGGCTCTGCTGGTCTGCCCTAGTTGATGATTTGGACATGCTTTGAATCTATG 1140
QY 1303 GTTAATAGCTTCAAGATATCACTCGGTCGCAACCGTGCATTTACCGATTTGAAGTATACAGGA 1362
    || || || || || || || || || || || || || || || || || || || || ||
Db 1141 GTTAGAGTGTTCGAACATCTCACTGTTGGTCACTGACTCAGTATGGCATGAACACATGCGG 1200
QY 1363 GCGCTTCCCAATATCTCAATATGGGAGT 1391
    || || || || || || || || || || || || || || || || || || || || ||
Db 1201 GCATTGCGCAACGTTTGCACAGCGCGT 1229
```

RESULT 13

AAQ50575

ID AAQ50575 standard: cDNA to mRNA: 1640 BP.

XX AC AAQ50575;

XX DT 24-MAY-1994 (first entry)

XX DE Asparaginylendopeptidase clone ASN-1.

XX KW Asparaginylendopeptidase; Canavalia ensiformis; seed:

XX KW L-asparagine; primer; PCR; protein fragmentation;

XX KW peptide synthesis; ss.

XX OS Canavalia ensiformis.

XX PN JP05276960-A.

XX PD 26-OCT-1993.

XX PF 07-AUG-1992; 92JP-0231602.

XX PR 07-FEB-1992; 92JP-0056023.

XX PA (SYUZ/) SYUZO T.

XX DR WPI; 1993-373587/47.

XX PT New gene for encoding asparaginyl endo-peptidase - comprises 8

XX PS specified DNA sequences

XX PS Disclosure; Page 26; 35pp; Japanese.

XX CC A gene encoding asparaginylendopeptidase is claimed.

XX CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

XX CC derived from a seed of Canavalia ensiformis which selectively

XX CC hydrolyses C-terminus amide bond of L-asparagine residue

XX CC (see AAM3033 and AAM3041).

XX CC The enzyme is useful for protein fragmentation and enzymatic

XX CC peptide synthesis.

XX CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used

XX CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,

XX CC by PCR.

XX SQ Sequence 1640 BP: 533 A; 288 C; 383 G; 436 T; 0 other;

Query Match 20.7%; Score 323.6; DB 14; Length 1640;

Best Local Similarity 55.8%; Pred. No. 3.1e-73;

Matches 664; Conservative 0; Mismatches 514; Indels 12; Gaps 2;

QY 208 TATTATAACTACAGGCATCAGGCTGACATATGCCACCGCTATCAGATCTATCCGAAAGGC 267

```

Db 4 TATGGTAATTTATCGGCATCAAGCGGATGTGTGCATCTTACCAGTTGCTGATAAAAGGT 63
QY 268 GTTTAAAGATGAAACATCATTTGCTTTATGTATGATCATATCGGTTTCTCGGAG 327
Db 64 GGAGTCNAAGAAGAACATTTGGTGTGTTATGTATGATGATATAGCTTATAACGCCATG 123
QY 328 AATCCTAGCGCTGGAGTTATCATTAATAAACAGATGAGAGATGTTTATAAAGGAGTT 387
Db 124 AATCCAGAGCCCGAGTCATCATCAACCATCTCAGGGCCAGAGGTGTTATGCTGTGTA 183
QY 388 CTTAAGGACTACACTAAGAAGCTGTTAATGTTCAAAACTTCTACATGTTACTTGGGA 447
Db 184 CTTAAGGATTACACCGGTGAGGAGCTAACAACCTTGAGAACCTATATGCTGTCATTTGG 243
QY 448 AATGAAGTGGCTTCACAGCAGGAAATGGCAAGTTGTTGAAAGTGTGCTCTAATGATAAT 507
Db 244 GACAAGAGTAAAGTTAAAGTGGAGTGGCAAGGTGATCAACAGCAATCCGGAGGATAGG 303
QY 508 ATCTTCATCTATTATGCTGACCATGGAGCTCTGGCTTAATAGCGATGCCCACTGTTGAT 567
Db 304 ATATTTATATCTACTCCGATCATGGAGTCCCGGAGTTCTTGGGATGCCAAACGCACCA 363
QY 568 GAAGTTATGCAAAAGATTTCAATGAAGTCTTGGAGAGATGATGATAGAGAAAATATAC 627
Db 364 TTCGTTTATGCCATGGATTTTATGATGTTTGAAGAAGAAACATGCAAGTGGAGGCTAC 423
QY 628 AACAAAGATGTGATCTATGTTGAAGCATGTGAATCAGAGATGATGTTTGAAGGATTTTA 687
Db 424 AAGAGATGTTTATACATAGAGCTGTGAGAGTGGGAGCATATTTGAGGGTATTAATG 483
QY 688 AAGAAAATCTCAACATATACGAGTACTGCTGCTAATTTCTAAAGAGACAGCTGGGA 747
Db 484 CCAAGGATCGAATATTTATGTGACAACCTGCGTCAAAATGCACAAGAGAACATTTTGA 543
QY 748 GTTTACTGCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 807
Db 544 ACTATTGCTCTGGGATGAATCTCTCTCCACGAGAGTACGTAACCTGCTCTCTCTCTCT 603
QY 808 ACATTTAGCATCTCTGCTGTTGAGGACAGTACCTTTCATGACATGACCAAGAGACTTG 867
Db 604 TTATACAGCTGTTCTGGATGGAGATAGTGAGACTCACAATCTAAAGGGAACCGTA 663
QY 868 GAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTGATGTACAGAGA-----CT 921
Db 664 CAACAGCAATACCAGTCGGTGAAGGAAACGGACTTCAAAATCTAACAGCTATAGTTTGGT 723
QY 922 TCTCATGTATGCGGTTTCGGAACAGAGAAGATGCTTAAAGATATCTTCTCTCTTACATT 981
Db 724 TCTCATGTATGCAATACGCTGACACTAACATTAATCTGCTGAAAAGCTTTTACTTGTACC 783
QY 982 GGAAGAAATCTGAAAACGATAACTTCACTTTACGGAATCCTTTTCTCACCACATCTCT 1041
Db 784 GTTTTCATCTGCCACCGTGACTTTCCTCC-----ACACACGCGCACTTAGAGCT 837
QY 1042 AATTCTGGTGTGTTCAATCCGCGGATATTCCTCTGCTGATACCTCCAGAGAAAGATTCAA 1101
Db 838 AAAATGGAAGTTGTTAACCAAGAGAGATGCGAGCTTTTGTTCATGTGCAAAATGATCAG 897
QY 1102 AAAGCTCCATGGATCATCTTGAAGCAAGCAAGCTCAGAGAAATGCTTGCAGAAAG 1161
Db 898 AGATCAAAACCATCAACCGGAAAAGAGACTCACATCTCTGGAACAGATTACAGAGACAGT 957
QY 1162 AATCATAGGAAAACAAATCGATCAGAGCATTAACAGACATTTCTGCGGCTTTCAGTTAAACAA 1221
Db 958 AAGCATAGGAATCATTTGGATGGCAGTGTGGAATTTGATGGAGTTTGTGTAGGACCA 1017
QY 1222 ACCAATGCTTAAATCTTCTTAACTTCCACAAGAACAAACAGAGACAGCTCTTTAGACGAT 1281
Db 1018 GGAAGAAAGTTCTTCGGTCTACATTTCCGTGAGGCTCTCTGCTGCGCCCTAGTTGATGAT 1077
QY 1282 TGGGATGCTTCAAGACTCTAGTTAATAGCTTCAAGAAATCATCTGCGGTGCAAGGTCAT 1341
Db 1341 TGGGATGCTTCAAGACTCTAGTTAATAGCTTCAAGAAATCATCTGCGGTGCAAGGTCAT 1341
```

```

Db 1078 TGGACATGCTTGAATCTATGTTAGAGTGTTCGAACTCAGTGTGGTCACTGACTCAG 1137
QY 1342 TAGCGATTGAAGTATACAGAGCGCTTCCCAATATCTGCAATATCTGCAATATGGGAGT 1391
Db 1138 TATGGCATGAACACATGCGGCGATTCGGCAACGCTTTGCAACACGCGCGT 1187

RESULT 14
AAQ50564
ID AAQ50564 standard; cdna to mRNA; 1152 BP.
XX AAQ50564;
AC AAQ50564;
XX 24-MAY-1994 (first entry)
XX Asparaginylendopeptidase sequence (6).
XX Asparaginylendopeptidase; Canavalia ensiformis; seed;
XX L-asparagine; primer; PCR; protein fragmentation;
XX peptide synthesis; ss.
XX Canavalia ensiformis.
XX JF05276960-A.
XX 26-OCT-1993.
XX 07-AUG-1992; 92JP-0231602.
XX 07-FEB-1992; 92JP-0056023.
XX (SYUZ/) SYUZO T.
XX WPI; 1993-373587/47.
XX New gene for encoding asparaginyl endo-peptidase - comprises 8
XX specified DNA sequences
XX Claim 1; Page 14-15; 35pp; Japanese.
XX A gene encoding asparaginylendopeptidase is claimed.
XX 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
XX derived from a seed of Canavalia ensiformis which selectively
XX hydrolyses C-terminus amide bond of L-asparagine residue
XX (see AAR43033 and AAR43041).
XX The enzyme is useful for protein fragmentation and enzymatic
XX peptide synthesis.
XX The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
XX in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
XX by PCR.
XX SQ Sequence 1152 BP; 340 A; 210 C; 297 G; 305 T; 0 other;

Query Match 20.5%; Score 319.2; DB 14; Length 1152;
Best Local Similarity 56.3%; Pred. No. 3.7e-72;
Matches 644; Conservative 0; Mismatches 488; Indels 12; Gaps 2;

QY 169 GGCACAGATGGGCTGTTTAGTGTGGATCAATGAATATTAATACTACAGGCATCAG 228
Db 7 GTTACCGGTGGGCGGTTCTGCTGGTTCATCAACGGTTATGGAATATTAGGCATCA 66
QY 229 GCTGACATATGCCACGGTATCAGATATCTCCGAAAAGCGGTTTAAAGATGAAAACATC 288
Db 67 GCGGATGTGGCATGCTTACCAGTTCTGCTAATAAAGGTGGAGTCAAGAAGAAACATT 126
QY 289 ATTGTGTTTATGATGATGATATCGCGTTTCTCGGAGATCCTAGCCCTGGAGTTATC 348
Db 127 GTGGTGTGTTATGATGATGATATAGCTTATAACCCCATGAATCCAGACCCGCGAGTCATC 186
QY 349 ATTAATAAACAGATGGAGAAGATGTTTATAAAGAGGTTCCTCAAGGACTACACTTAAAGAA 408
Db 187 ATCAACCATCTCAGGGGCCAGACGCTGATGCTGTACCTAAGGATTACACCGGTGAG 245
```


649	QY	GAAGCATGTGAATCAGCAGAGTATGTTTGAAGCGATTATTAAGAAAAAATCTCAACATATAC	708
487	Db	GAAGCTTTGTGAGAGTGGGAGCATATTTGAGGGTATTAATGCCAAGGATCTGAATATTTAT	546
709	QY	GCAGTCAGCTGCTCTAATCTTAAGAGAGACAGCTGGGAGTTTACTGTCTCCTGAGTCATAT	768
547	Db	GTGACAACCTGCGTCAATGCAAGAGAGACAGTTTGGAACTTATTTGCTCTGGATGAAT	606
769	QY	CCTCTCTCCCTTCTTGAGATTGGAACCTTGCTCGGCGATACATTTAGCATCTCTTGGCTT	828
607	Db	CCTCTCCACCAAGACAGTACGTAACTTGCTCGGGGATTTATACAGCGTTCTTGGATG	666
829	QY	GAGGACAGTACCTTCATGACATGAGCAAGACAGCTTTGGAGCAACAATACACAGTTGTA	888
667	Db	GAAGATAGTGAGACTCACAACTCTAAAAGGGAACGGTACACAGCAAAATACCAAGTCGGTA	726
889	QY	AAGAGAAGAGTAGGATCTGATGTACCAGAGA-----CTTCTCATGCTATGCGGTTTCGGA	942
727	Db	AGGAACGGACTTCAATTTCTAACAGCTATAGGTTTGGTTTCTCATGTGATGCAATACGGT	786
943	QY	ACAGAGAAGATGCTTAAAGATTATCTTCTCTTACATTTGGAGAAATCCTGAAAAACGAT	1002
787	Db	GACACTAACATTACTGCTGAAAAGCCTTACTTGTAACCATGGTTTGTGATCCTGCCACCGTG	846
1003	QY	AACTTCACT	1011
847	Db	AACTTTCCCT	855

Search completed: May 29, 2003, 15:24:41
Job time : 402 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:17:54 ; Search time 100 Seconds
(without alignments)
4784.157 Million cell updates/sec

Title: US-09-934-066-1
Perfect score: 1560
Sequence: 1 ctcaagaatcagattca.....gaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198.8	12.7	1936	2	US-09-139-424-3
2	198.8	12.7	2030	3	US-08-706-216-3
3	190.8	12.2	1855	2	US-08-928-613-1
4	190.8	12.2	1855	3	US-09-193-524-1
5	52	3.3	7218	1	US-08-232-463-14
6	45	2.9	1492	4	US-08-745-995A-25
7	45	2.9	1492	4	US-08-745-995A-27
8	41.4	2.7	168575	4	US-09-426-290-1
9	40.8	2.6	206	2	US-08-928-613-22
10	40.8	2.6	206	3	US-09-193-524-22
11	40	2.6	859	4	US-09-247-373B-47
12	39.6	2.5	8920	2	US-08-446-855A-1
13	39.6	2.5	8920	4	US-09-150-741-1
14	38.4	2.5	544	4	US-09-280-116-247
15	38.4	2.5	6474	4	US-08-961-527-155
16	38.2	2.4	72928	3	US-09-009-913-1
17	37.8	2.4	155	2	US-08-928-613-5
18	37.8	2.4	155	3	US-09-193-524-5
19	37.4	2.4	5020	4	US-08-961-527-142
20	37.2	2.4	219	2	US-08-928-613-11
21	37.2	2.4	219	3	US-09-193-524-11
22	37.2	2.4	220	2	US-08-928-613-15
23	37.2	2.4	220	3	US-09-193-524-15
24	36.8	2.4	591	3	US-08-882-501-25
25	36.2	2.3	403	4	US-09-385-982-29
26	36.2	2.3	5852	1	US-07-867-106-2
27	36	2.3	1530	4	US-09-134-001C-159

c	28	35.6	2.3	1388	4	US-09-144-776B-9	Sequence 9, Appli
c	29	35.6	2.3	1712	4	US-09-144-776B-5	Sequence 5, Appli
c	30	35.6	2.3	1712	4	US-09-144-776B-7	Sequence 7, Appli
c	31	35.6	2.3	5852	1	US-07-867-106-2	Sequence 2, Appli
c	32	35.2	2.3	929	3	US-08-860-368B-4	Sequence 4, Appli
c	33	35.2	2.3	1158	3	US-08-860-368B-1	Sequence 1, Appli
c	34	35	2.2	4118	1	US-08-119-125A-3	Sequence 3, Appli
c	35	34.8	2.2	9542	4	US-08-968-685A-9	Sequence 9, Appli
c	36	34.8	2.2	17656	4	US-09-433-579-3	Sequence 24, Appli
c	37	34.4	2.2	591	3	US-08-882-501-24	Sequence 24, Appli
c	38	34.4	2.2	2415	4	US-09-134-001C-2381	Sequence 2381, Ap
c	39	34.2	2.2	1308	4	US-09-150-213-1	Sequence 1, Appli
c	40	34	2.2	199	2	US-08-928-613-18	Sequence 18, Appli
c	41	34	2.2	199	3	US-09-193-524-18	Sequence 18, Appli
c	42	34	2.2	1609	4	US-09-312-285-1	Sequence 1, Appli
c	43	34	2.2	1609	4	US-09-728-764-1	Sequence 1, Appli
c	44	34	2.2	1720	4	US-09-227-357-139	Sequence 139, App
c	45	34	2.2	1972	4	US-09-370-098-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-139-424-3
; Sequence 3, Application US/09139424
; Patent No. 5985832
; GENERAL INFORMATION:
; APPLICANT: Roodman, G. David
; APPLICANT: Reddy, Sakamuri V.
; APPLICANT: Choi, Sun-Jin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE FOR
; TITLE OF INVENTION: OSTEOCLAST INHIBITORY FACTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,424
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/772,441
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSK:295
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1936 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..1434
; US-09-139-424-3

Query Match 12.7% Score 198.8; DB 2; Length 1936;
Best Local Similarity 59.1%; Pred. No. 8.9e-46;

Matches 429; Conservative 0; Mismatches 267; Indels 30; Gaps 4;

Qy 168 AGGCACAGATGCGCTGTTTCTGCTGATCAATGAATATTAATACTACAGCATCA 227
 Db 213 AGGCACAGATGCGCTGTTTCTGCTGATCAATGAATATTAATACTACAGCATCA 272
 Qy 228 GGCTGATATGCGCGGTATCAGATCTCGAAGCGGTTTAAAGATGAAACAT 287
 Db 273 GGCGAGCGGTGCGCTACAGATCAATCAGCGCAATGGGATCTCTGAGCAACAGAT 332
 Qy 288 CATTTGTTTATGATGATATCCGTTTCTCGGAGATCTCTAGGCTTGGAGTTAT 347
 Db 333 CGTTGATGATGATGATGATCTCTGAGGATCTCTAGGCTTGGAGTTAT 392
 Qy 348 CATTAATAACAGATGAGGATGTTTATAAGGAGTTCTTAAGGATCTACATTAAGA 407
 Db 393 GATCAACAGGCGCAATGGCAGATCTCTAGGAGTCCCGAAGGATCTACATGGA 452
 Qy 408 AGCTGTTAATGTTCAAACTTCTACATGTTTACTTGGAAATGAAAGTGGCGTCACAGG 467
 Db 453 GGATGTTTACCCCAAAATTTCTGCTGTTGAGAGCGGATGCGAAGCAGTGAAGG 512
 Qy 468 ---AGGAATGCGCAAGTTGTGAAAGTGTCTTAATGATATATCTTATCTATTATGC 524
 Db 513 CATAGGATCGGCAAGTCTGAAGATGCGCCCGAGGATCAGGTTCATTCTTCTCAC 572
 Qy 525 TGACCATGAGCTCTGCTGTTTATACGATGCGCCCTGCTGATGAGTATGCGCAAGA 584
 Db 573 TGACCATGAGCTCTGCTGTTTATACGATGCGCCCTGCTGATGAGTATGCGCAAGA 629
 Qy 585 TTTCAATGAGTCTGCGAAGATGATTAAGAGATGATTAAGAGATGATTAAGAGATGAT 644
 Db 630 CTTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 689
 Qy 645 TGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 704
 Db 690 CATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
 Qy 705 ATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 764
 Db 747 TTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 805
 Qy 765 ATATCT 824
 Db 806 ---AGAGTCT 845
 Qy 825 GCTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 884
 Db 846 GATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 905
 Qy 885 TGTAAA 890
 Db 906 GGTAAA 911

RESULT 2

US-08-706-216-3
 : Sequence 3, Application US/08706216
 : Patent No. 614098
 : GENERAL INFORMATION:
 : APPLICANT: Balasubramanian, Sriram
 : APPLICANT: Ford, John
 : APPLICANT: Gozian, Daniel M.
 : APPLICANT: Zurewsky, Gerard
 : TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : STREET: DNAX Research Institute
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: USA
 : ZIP: 94304-1104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/706,216
 FILING DATE: 30-AUG-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Ching, Edwin P.
 REGISTRATION NUMBER: 34,090
 REFERENCE/DOCKET NUMBER: DX0613
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-852-9196
 TELEFAX: 415-496-1200
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2030 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 183..1484
 US-08-706-216-3

Query Match 12.7%; Score 198.8; DB 3; Length 2030;

Best Local Similarity 59.1%; Pred. No. 9.1e-46;
 Matches 429; Conservative 0; Mismatches 267; Indels 30; Gaps 4;

Qy 168 AGGCACAGATGCGCTGTTTCTGCTGATCAATGAATATTAATACTACAGCATCA 227
 Db 213 AGGCACAGATGCGCTGTTTCTGCTGATCAATGAATATTAATACTACAGCATCA 319
 Qy 228 GGCTGATATGCGCGGTATCAGATCTCGAAGCGGTTTAAAGATGAAACAT 287
 Db 320 GGCGAGCGGTGCGCTACAGATCAATCAGCGCAATGGGATCTCTGAGCAACAGAT 379
 Qy 288 CATTTGTTTATGATGATATCCGTTTCTCGGAGATCTCTAGGCTTGGAGTTAT 347
 Db 380 CGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 439
 Qy 348 CATTAATAACAGATGAGGATGTTTATAAGGATGATTAAGGATCTACATTAAGA 407
 Db 440 GATCAACAGGCGCAATGGCAGATCTCTAGGAGTCCCGAAGGATCTACATGGA 499
 Qy 408 AGCTGTTAATGTTCAAACTTCTACATGTTTACTTGGAAATGAAAGTGGCGTCACAGG 467
 Db 500 GGATGTTTACCCCAAAATTTCTCTGCTGTTGAGAGCGGATGCGAAGCAGTCAAGG 559
 Qy 468 ---AGGAATGCGCAAGTTGTGAAAGTGTCTTAATGATATATCTTATCTATTATGC 524
 Db 560 CATAGGATCGGCAAGTCTGAAGATGCGCCCGAGGATCAGGTTCATTCTTCTCAC 619
 Qy 525 TGACCATGAGCTCTGCTGTTTATACGATGCGCCCTGCTGATGAGTATGCGCAAGA 584
 Db 620 TGACCATGAGCTCTGCTGTTTATACGATGCGCCCTGCTGATGAGTATGCGCAAGA 676
 Qy 585 TTTCAATGAGTCTTGGAGATGATTAAGAGATGATTAAGAGATGATTAAGAGATGAT 644
 Db 677 CTTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736
 Qy 645 TGTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 704
 Db 737 CATTTGAGCTGTTGATGCTGGGTCTCATGATGAA---CCACCTGCGGATGATCAATGT 793
 Qy 705 ATACGAGTACTGCTGCTTAAATCTTAAGAGAGAGCTGGGAGTCTTACTGCTGCTG 764
 Db 794 TTATGCAACTACTGCTGCGCAAGCCCGAGAGTCTGCTACGCTCTGTTACTATGATGAGA- 852
 Qy 765 ATATCT 824

[illegible]

RESULT 3
US-08-928-613-1
; Sequence 1, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeans, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

431 TATAGGATCCCGAAAGTCTGAAGAGTGGTCCCAGGATCAGCGTGTTCATTTATTTTCAC 490

525 TGACCATGGAGCTCCTCGCTTAATAGCGATGCCCACTGGGTGATGAAAGTTATGGCAAAAGA 584

491 TGACCATGGATCTTCTGGAATACTGGTTCATTTCCCA---ATGAAGATCTTTCATGTAAGA 547

585 TTTCAATCAAGTCTTGGAGAAGATGCATAAGAGAGAAAAAATACACAAGATGGTGATCTA 644

548 CCGTATTAAAGACCAACCATACATTTTCAAAAAACAATGTACCGAAAGATGGTGTCTTA 607

645 TGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAAAATCTCAACAT 704

608 CATTGAGCGCTGTGAGTCTGGTCCATGATGAACCACTGCCGGAATAACA--TCAATGT 664

705 ATACCCAGTACTGCTGCTAATCTTAAAGAGAGAGCTGGGAGTTTACTGCTCTGAGTC 764

665 TTATCAACTACTGCTGCCAACCCAGAGAGTCTCTACGCCCTGTACTATGATGAGA- 723

765 ATATCTCTCTCTCTCTCTGAGATTGGAACCTTGCTCTGGGATACATTTAGCATCTCTTG 824

724 -----AGAGTCCAGTCTCTGGGGACTGGTACAGCGTCAACTG 763

825 GCTTGAGGACAGTGACCTTTCATGACATGAGCAAGAGAGACTTTGGAGCAACAATACCACT 884

- 764 GATGAAGACTCGGACGTGGAAGATCTGACTAAGAGAGACCTTGCACAAGCAGTACCACT 823

885 TGATAA 890

824 GGATAA 829

RESULT 4

US-09-193-524-1

; Sequence 1, Application US/09193524

; Patent No. 6007997

; GENERAL INFORMATION:

; APPLICANT: Diep, Dinh

; APPLICANT: Braxton, Scott M.

; APPLICANT: Deleane, Angelo M.

; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/193,524

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/928,613

; FILING DATE:

```

1 COUNTRY: USA
2
3 ZIP: 94304
4
5 COMPUTER READABLE FORM:
6
7 MEDIUM TYPE: Floppy disk
8
9 COMPUTER: IBM PC compatible
10
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12
13 SOFTWARE: Patentin Release #1.0, Version #1.30
14
15 CURRENT APPLICATION DATA:
16
17 APPLICATION NUMBER: US/09/193,524
18
19 FILING DATE:
20
21 CLASSIFICATION:
22
23 PRIOR APPLICATION DATA:
24
25 APPLICATION NUMBER: 08/928,613
26
27 FILING DATE:
28

```

[illegible]


```
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berigling Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1998-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
; US-09-426-290-1

Query Match      2.78; Score 41.4; DB 4; Length 168575;
Best Local Similarity 53.4%; Pred. No. 0.57;
Matches 87; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1397 TGAAGCAACCTGTTTCAGCCATTGAACACCTTTGTCGATGTAATGATTTGCAAAACAAT 1456
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1457 GTGATATTCGACTTTAAATAATCAAGTTAATTTCAATAAACTCGATGTCAGAGATGCT 1516
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14868 TCAATGTCAGCTAAAAAGTTAAGACTGTGATTTTAAATAAATACTAGATTTAGAAATAAA 148527
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1517 TGGTTCATGATCTACTTTTACATGAAAAAATAAAAAAAAAAAAA 1559
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148528 TCAAAATTGAAATGACATTTATTAACCTTAAAAATAACAAAAAAA 148570
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
; US-09-426-290-1
; Sequence 22, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleageane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
```

```
US-09-928-613-22
; APPLICATION NUMBER: US/08/928,613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 206 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Kidney
; CLONE: 195541
; US-09-928-613-22

Query Match      2.68; Score 40.8; DB 2; Length 206;
Best Local Similarity 59.5%; Pred. No. 0.034;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 564 TGATGAAGTTATGCGCAAAAGATTTCATGAAGTCTTGGAGAAAGATGCATAAGAGAAAAA 623
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 624 ATACAACAAGATGCTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTT 674
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 GTACCGAAAGATGCTGTTCTACATTNAGCGCTGTNAGTCTGGTCCATGTT 125
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
; US-09-193-524-22
; Sequence 22, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleageane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
```



```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 206 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; IMMEDIATE SOURCE:
; LIBRARY: Kidney
; CLONE: 195541
; US-09-193-524-22

Query Match      2.6%; Score 40.8; DB 3; Length 206;
Best Local Similarity 59.5%; Pred. No. 0.034;
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 564 TGATGAAGTTATGCGAAGATTTCAATGAAGTCTTGGAGAGATGCAATGAAGAGAAAAA 623
    ||| || || || || || || || || || || || || || || || || || || ||
Db 15 TGAAGATCTTCATGTAAGGACCTGANTGAGACCATCCATTACATGTACAAACACAAAA 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 624 ATACAACAGATGTCATCTATCTGAAGCATGTCATGATGAGGATGTT 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 GTACCGAAGATGGTGTCTACATTNAGGCGCTGTCGTCATGTT 125

RESULT 11
US-09-247-373B-47
; Sequence 47, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 859
; TYPE: DNA
; ORGANISM: SOYBEAN
; US-09-247-373B-47

Query Match      2.6%; Score 40; DB 4; Length 859;
Best Local Similarity 57.0%; Pred. No. 0.11;
Matches 73; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 527 ACCATGAGCTCTGCTTAATGCGATGCCACTGGTGATGATGCGAAGATT 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 AGCTGGGAAGCTGGATGGTGGTACGATGCTACCGCGGGGAAGAACAGGAAGCGTG 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 587 TCAATGAAGCTCTTGGAGAGATGCATAAGAGAAAAAATACAAAGATGGTGATCG 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 TGAAGGAAGCCATAGAATGATGAGAAATAGAGAGAGATTAAGGGAAGAAATTTT 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 647 TTGAAGCA 654
    ||| ||| |||
Db 442 TTGGAGGA 449

RESULT 12
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
```

```
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match      2.5%; Score 39.6; DB 2; Length 8920;
Best Local Similarity 54.9%; Pred. No. 0.44;
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1419 TGAACAGCTTGTTCGATGTAATGATTGCAAAACAAATGATTCGACTTTAAAAATA 1478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TTAATAAGTTTTATATTTCAAGTAATTTTATAACAAATGAACACACACATATATATA 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1479 TCAAAAGTTAATTTCAATAAAACTCGATGTAGAGATGTTGGTTCATGACTACTTTTAC 1538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1539 ATGAAAAAATAAAAAAAAAAAAAA 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 ATATAAGAAAAAATAAAAAAAAAA 256

RESULT 13
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; PATENT NO. 6183996
; FILE REFERENCE:
; SYNTHETASE II
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match      2.5%; Score 39.6; DB 4; Length 8920;
Best Local Similarity 54.9%; Pred. NO. 0.44; 64; Indels 0; Gaps 0;
Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1419 TGAACAGCTGTTTCGATGTAATGTTGCAAAAACATGTGATATTCGACTTTAAATA 1478
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 397 TTAATAAGTTTATATTTCAAGTAATTTTATAACAAATGAACACACACATATATATA 338
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1479 TCAAGTTTAAATTCATATAAAATCGATGATGAGATGTTGGTTTCATGATACACTTTTAC 1538
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 337 TATATATATATATATATATATATATATATATAATAAACTTAAATGTTTAAATAA 278
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1539 ATGAAAAAATAAAAAAAAAAAAA 1560
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 277 ATATAAGAAAAAATAAAAAA 256
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-280-116-247
; Sequence 247, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: hemoglobinase
; NAME/KEY: misc_feature
; LOCATION: (1)..(544)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-247

Query Match      2.5%; Score 38.4; DB 4; Length 544;
Best Local Similarity 60.6%; Pred. NO. 0.25;
Matches 63; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 226 CAGCGTGACATATGCCACGCGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAAC 285
Db      ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 64 CAGCGACGCGTGCCATCGCTACCAGATCATTCACCGCAATGGGATTCCTGACGACAG 123
Db      ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 286 ATCATTTGTTTATGATGATGATCGCGTTTCTCGGAGAA 329
Db      ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 124 ATCGTTGTGATGATGTACGATGACATTCGTTACTCTGAAGAGTA 167
Db      ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-08-961-527-155/C
; Sequence 155, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
```

```
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-155

Query Match      2.5%; Score 38.4; DB 4; Length 6474;
Best Local Similarity 56.2%; Pred. NO. 0.82; 56; Indels 0; Gaps 0;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 1427 CTTGTTTCGATGTAATGTTGCAAAACATGTCATATTCGACTTTAAAAATATCAAGTT 1486
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 4572 CTTTTTGAATATATGTTATATATATATATATATATATATATATATATATATATAT 4513
Db      ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1487 AATTTCAATAAAACTCGATGTAGAGATGTTGGTTTCATGATACACTACTTTTACATGAAAA 1546
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 4512 ATGTTCAATTAATAATTTAGAAAAGTGAGAAAACATTTGTTGTATATATATAAATCAATA 4453
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1547 AAAAAAA 1554
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 4452 TAAAAAA 4445
Db      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: May 29, 2003, 17:10:05
Job time : 212 secs
```


QY 385 GTTCCCTA----- 391
Db 2461 GTTCCCTAAGGTTCTTATTCTACTCTCTTTTGTGCGTTATTCTACGTTGAATTCATTTAC 2520
QY 392 -----AGGACTACACTAAGAAGCTGTGA 415
Db 2521 ATATATATATCAAGTTTGTGTTATTTGGTTAGGCTACACTAAGAAGCTGTGA 2580
QY 416 ATGTTCAAACTTCTACAATGTTTACTTGGAAATGAAAGTGGCGTCACAGAGGAGAAATG 475
Db 2581 ATGTTCAAACTTCTACAATGTTTACTTGGAAATGAAAGTGGCGTCACAGAGGAGAAATG 2640
QY 476 GCAAAGTTGGAAGTGGTCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAG 535
Db 2641 GCAAAGTTGGAAGTGGTCTTAATGATAATATCTTCATCTATTATGCTGACCATGGAG 2700
QY 536 CTCTCGCTTAAT----- 548
Db 2701 CTCTCGCTTAATAGGTTTCTTAATTTATGAAATTTATACGTACCATCAATTCATATC 2760
QY 549 -----AGCGATGCCCACTG 562
Db 2761 TATAATAAAGATTTTCTTGATACACTACGAACCGCGATTTTCTCAGCGATGCCCACTG 2820
QY 563 GTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGGAGAAGATGCAATAAGAGAAAA 622
Db 2821 GTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTGGAGAAGATGCAATAAGAGAAAA 2880
QY 623 AATACACAAGAT----- 635
Db 2881 AATACACAAGATGGTATATACTCAACCATTCGTTACCTAGCTTTTATACATATGTGTT 2940
QY 636 -----GGTGATCTATGTTGAAGCATG 656
Db 2941 TGTTTTGATCTCTATGTTGTTTGGATGTTTGGATGTTTGGATGTTTGGATGTTTGGATGTT 3000
QY 657 TGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTCAACATATACGAGTGAC 716
Db 3001 TGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTCAACATATACGAGTGAC 3060
QY 717 TGCTGCTAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCCTCTCC 776
Db 3061 TGCTGCTAATCTTAAAGAGAGCAGCTGGGAGTTTACTGTCCTGAGTCATATCCTCTCC 3120
QY 777 TCCTCTGAGATTGGAATTTGCTCGCGGATACATTTAGCATCTCTTGGCTGA----- 830
Db 3121 TCCTCTGAGATTGGAATTTGCTCGCGGATACATTTAGCATCTCTTGGCTGAGGACAG 3180
QY 831 ----- 830
Db 3181 GTACTGCAAAACAAAAAGATTCAATCCTTATGGACTATTCGAATGATTGATTGTTCTT 3240
QY 831 -----GGACAGTACCTTCATG 847
Db 3241 GAGAAATATTGTTTCATTTGTTCTATGTTTGTGTTGTTGGAGACAGTACCTTCATG 3300
QY 848 ACATGAGCAAGAGACTTTGGAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTG 907
Db 3301 ACATGAGCAAGAGACTTTGGAGCAACAATACCACGTTGTAAGAGAGAGTAGGATCTG 3360
QY 908 ATGTACAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAAGATGCTTTAAAGATTATC 967
Db 3361 ATGTACAGAGACTTCTCATGTATGCGGTTTTCGGAACAGAGAAGATGCTTTAAAGATTATC 3420
QY 968 TTTCCCTTACATGGAAGAAATCCTGAAAACGATACCTTTCACCGGAATCCTTTT 1027
Db 3421 TTTCCCTTACATGGAAGAAATCCTGAAAACGATACCTTTCACCGGAATCCTTTT 3480
QY 1028 CCTCACCATACTTAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTCTATACCTCC 1087
Db 3481 CCTCACCATACTTAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTCTATACCTCC 3540

QY 1088 AGAGAA----- 1093
Db 3541 AGAGAAAGGTGAGCTTTTTCGGGTTTTTTTGATCATTTTAAACGAAAGAGTTTTTTCAGCAT 3600
QY 1094 -----AGATTCAAAAGCTCCAATGGGATCACTTTGAAAGCAA 1130
Db 3601 GTTTTAAATGTTTATTCATCTCTTAGATTCAAAAGCTCCAATGGGATCACTTTGAAAGCAA 3660
QY 1131 AGAAGCTCAGAGAAAATGCTTGACGAAAAGATCATAGAAACAATCGATCAGAGCAT 1190
Db 3661 AGAAGCTCAGAGAAAATGCTTGACGAAAAGATCATAGAAACAATCGATCAGAGCAT 3720
QY 1191 TACAGACATTTCTCGGCTTTTCAGTTAAACAACAATGCTTAAATCTCTTAACTTCCAC 1250
Db 3721 TACAGACATTTCTCGGCTTTTCAGTTAAACAACAATGCTTAAATCTCTTAACTTCCAC 3780
QY 1251 AGAACAACAGGACAGCTCTTGACAGCATTCGGATTCGTTCAAGACTCTA----- 1302
Db 3781 AGAACAACAGGACAGCTCTTGACAGCATTCGGATTCGTTCAAGACTCTTAGTAACAAA 3840
QY 1303 ----- 1302
Db 3841 CCACATCTCAAACTTTGTTACTTTGTTCTTACCGCAACACCATTCGATTATTACTAAACC 3900
QY 1303 -----GTTAATAGCTTCAAGATCACTCGGTCGAACGGT 1337
Db 3901 AGTGTATATCGAATGAAAAATCGCAGGTTTATAGCTTCAAGAAATCACTCGGTCGAACGGT 3960
QY 1338 GCATTACGGATTCAAGTATACAGGAGCGCTTGCCTCAATATCTGCAATATGGAGTGGATGT 1397
Db 3961 GCATTACGGATTCAAGTATACAGGAGCGCTTGCCTCAATATCTGCAATATGGAGTGGATGT 4020
QY 1398 GAAGCAAACTGTTTTCAGCCATTGAACAAGCTTCTGCTGATTAATGATTTGCAAAACAATG 1457
Db 4021 GAAGCAAACTGTTTTCAGCCATTGAACAAGCTTCTGCTGATTAATGATTTGCAAAACAATG 4080
QY 1458 TGATATTCGACTTTTAAATAATCAAAAGTTAATTTCAATAAACTCGATGATAGATGGTT 1517
Db 4081 TGATATTCGACTTTTAAATAATCAAAAGTTAATTTCAATAAACTCGATGATAGATGGTT 4140
QY 1518 GGTTCATGACTACTCTTTTACATGAAAA 1545
Db 4141 GGTTCATGACTACTCTTTTACATGAAAA 4168

RESULT 3

US-09-938-842A-1485
; Sequence 1485, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1485
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1485

Query Match 29.2% Score 454.8 DB 9 Length 1485

QY 885 TGTAAA 890
DB 951 GGTAAG 956

RESULT 7
US-09-878-574-3641
Sequence 3641 Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878-574
PRIORITY FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
NUMBER OF SEQ ID NOS: 1999-06-14
SEQ ID NO 3641
LENGTH: 393
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-007-Q1-B1-F5
US-09-878-574-3641

Query Match 12.3% Score 192.6; DB 10; Length 393;
Best Local Similarity 68.3%; Pred. No. 5e-40;
Matches 267; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 372 TGTATTATAAGGAGTTCTCAAGGACTACACTAAAGAAGCTGTTTAAATGTTCAAACACTTCTA 431
DB 1 TGTATTATAAGGAGTTCTCAAGGAGTTACACGGCGGAAGATGTTTACTGTTCGATAAAGCTTTT 60

QY 432 CAATCTGTAGTTTGGAAATGAAGTGCGCTCACAGGAGGAAATGCGCAAAAGTTGTGAAAAAG 491
DB 61 TGCTCTGTTTTACTTGGAAATAAGTCAGGACACTGACTGTGGCAGTGGGAAGGTTTGTGCAGAC 120

QY 492 TGGTCCCTAATGATAATATPTTTCATCTATTATGCTGACCATTGGAGCTCCTGGCTTAATAGC 551
DB 121 TGGTCCCTGATGATCATATATTTGTATATCTATATGACCTGGAGGTCTGGGGTGTCTCGG 180

QY 552 GATGCCCACTGGTGATGAAGTTATGGCAAAGATTTCAATGAAGTCTTGGAGAAGATGCA 611
DB 181 GATGCTCTGCTGCTCTACTTATACGGGGATGATCTGATTGAAGTCTTGAAGANAAGCA 240

QY 612 TAAGAAAAAATAACAACAGATGGTGATCTATGTTGAAGCATGTAATCAGGAAGTAT 671
DB 241 TCCTTCTGGACATATAAACCCTAGTATTTATCTGGAGGCAATGTAATCTGGGAGTAT 300

QY 672 GTTTGAGGGATTTTAAAGAAATCTCACATATACCGAGTGCCTGCTTAATTTCTAA 731
DB 301 CTTTGAGAGTCTTCTCTCGATATCATATTTATGCANCCCTGCTTCCANTGCAG 360

QY 732 AGAGAGAGCTGGGGAGTTTACTCTCTGAG 762
DB 361 ACAGAGTACTTGGGAACATATTCGCCCGG 391

RESULT 8
US-09-967-796-1
Sequence 1 Application US/09967796
Patent No. US2002015553A1
GENERAL INFORMATION:
APPLICANT: Diep, Dinh
APPLICANT: Braxton, Scott M.
Deleageane, Angelo M.
TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/967,796
FILING DATE: 28-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/449,422
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-0048 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: Adrenal
CLONE: 100877
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-967-796-1

Query Match 12.2%; Score 190.8; DB 9; Length 1855;
Best Local Similarity 58.4%; Pred. No. 3.5e-39;
Matches 424; Conservative 0; Mismatches 272; Indels 30; Gaps 4;
QY 168 AGGCACAGATGGGCTGTTTGGTGGATCAATGAATATATACTACAGCATCA 227
DB 131 AGGCAAGCAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 190
QY 228 GCGTGACATATGCCACGCGTATCAGATCTCCGAAAGGCGGTTTAAAGATGAACAT 287
DB 191 GGCAGCGGTGCCATGCCATGCCATGCCATGCCATGCCATGCCATGCCATGCCAT 250
QY 288 CATTGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 347
DB 251 CGTTGTGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310
QY 348 CATTAAATACAGATGAGAGATGTTTAAAGAGTTCCTTAAGGACTACATTAAGA 407
DB 311 GATCAACAGGCCCAATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 370
QY 408 AGCTGTTAATGTTCAAACTTCTACAATGCTTACTTGGAAATGAAAGTGGCTCAGG 467
DB 371 GGATGTTTACCCCAAAAATTTCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 430
QY 468 ---AGGAATGCAAAAGTTGTGAAAGTGGTCCCTAATGATAATATCTTCTATTTATGC 524
DB 431 TATAGATCCCGCAAGTCTGAGAGTGGTCCCGAGGATCAGCTGTTCAITTTATTCAC 490
QY 525 TGACCATGGAGTCTCTGCGTTAATAGCAGTCCCGACTGGTGGTGGTGGTGGTGGTGGT 584
DB 491 TGACCATGGATCTTCTGCAATACTGGTTTCCCA---ATGAAGATCTTCTATGTAAGA 547
QY 585 TTTCATCAAGTCTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 644
DB 548 CTTGATTAAGACCAACCCATTACATTTTCAAAAACAAAATGTACCGAAAGATGGTGTCTCA 607

QY 645 TGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAATCTCAACAT 704
DB 608 CATTGAGGCTGTGAGTCTGGTCCATGATGAACACCATGCCGATACACA---TCATGT 664
QY 705 ATACGCACTGACTGCTGCTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
DB 665 TTATGCAACTACTGCTGCCAACCCAGAGAGTCTGCTTACGCTGTTTACTATGATGAGA- 723
QY 765 ATATCCCTCTCTCTCTGAGATTGGAATTTGCTCGGCGATACATTTAGCATCTCTTG 824
DB 724 -----AGAGTCCACCTACCTTGGGGAGCTGGTACACGGCTCAACTG 763
QY 825 CTTTGAGGACAGTGAACCTTTCATGACATGAGCAAAAGAGACTTTTGGAGCAACATACCAGT 884
DB 764 GATGGAAGACTCGGAGCTGGGAAGATCTGACTAAAGAGACCCCTGCACAGCAGTACCACCT 823
QY 885 TGTAAG 890
DB 824 GGTAAA 829
RESULT 9
US-09-878-574-4049
; Sequence 4049, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(405)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-002-Q1-B1-C9
US-09-878-574-4049

Query Match 11.3%; Score 177; DB 10; Length 405;
Best Local Similarity 64.9%; Pred. No. 5.7e-36;
Matches 261; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 406 GAAGCTGTTAATGTTCAAACTTCTACAAATGTTTACTTGGAAATGAAAGTGGCGTCA 465
DB 2 GAAGATGTTACTGTTGTTAACTTTTCTGCTCTTACTTGGAAACAGTCGGCACTTACT 61
QY 466 GGAGAAATGCAAGTTGTGAAAGTGGTCCCTAATGATAATATCTTCTATTTATGCT 525
DB 62 GGTGGCAGTGGAGGTTGTGGACAGTGGAGCTGATGATCATATATTTGTATATCTACT 121
QY 526 GACCATGGAGCTCTGCTTAATAGCGATGCCACTTGGTGTGATGAAGTTTATGGCAAAAG 585
DB 122 GACCATGGAGCTCCAGGGTCTCGGGATGCTCTGCTGCTCTTACTTATATGCGGATCAT 181
QY 586 TTCAATGAAGTCTTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 645
DB 182 CTGATTGAAGTCTTGAAGAAAACATGCTTCTGGACGATATAAAACCTAGTATTTTAT 241
QY 646 GTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAAATCTCAACATA 705
DB 242 CTGGAGGCATGTGAATCTGGGAGTATCTTTGAAGGCTTCTTCCCGAAGATATCAATAT 301

Fri May 30 09:13:12 2003

us-09-934-066-1.rnpb

Page 10

Search completed: May 29, 2003, 18:21:46
Job time : 233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2003, 15:16:26 ; Search time 2083 seconds
(without alignments)
12129.121 Million cell updates/sec

Title: US-09-934-066-1

Perfect score: 1560

Sequence: 1 ctcaagaatcagattcaa.....gaaaaaaaaaaaaaaaaa 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pln.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_oth.*

26: em_gss_pro.*

27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c 1	585	37.5	585	10	AV559121
c 2	580.2	37.2	606	10	AV557684
c 3	561	36.0	561	10	AV559478
c 4	541	34.7	541	10	AV559545
c 5	535	34.3	535	10	AV561906
c 6	528.6	33.9	535	10	AV562667

7	524.2	33.6	538	10	AV555212
c 8	502.8	32.2	506	10	AV557191
c 9	492	31.5	574	10	AV831492
c 10	443	28.4	443	10	AV562308
c 11	437	28.0	437	10	AV563648
c 12	431	27.6	431	10	AV564659
c 13	428.2	27.4	462	10	AV559517
c 14	418	26.8	431	10	AV810306
c 15	411	26.3	411	10	AV819966
c 16	409.4	26.2	427	10	AV560592
c 17	409	26.2	437	10	AV826649
c 18	404.8	25.9	501	10	BE577730
c 19	403.8	25.9	416	10	AV800099
c 20	402.4	25.8	422	10	BE522129
c 21	399	25.6	399	10	AV806388
c 22	385.2	24.7	1981	11	AV104034
c 23	379.8	24.3	391	10	AV565648
c 24	378	24.2	400	10	AV799476
c 25	378	24.2	431	10	AV809370
c 26	367.2	23.5	404	10	AV559374
c 27	363.8	23.3	696	12	BG41524
c 28	354.6	22.7	745	14	BQ871135
c 29	354.6	22.7	760	14	BQ871979
c 30	347.8	22.3	356	10	BE524724
c 31	346.4	22.2	410	10	AV795983
c 32	345.8	22.2	357	10	AV564976
c 33	344.4	22.1	378	10	AV811879
c 34	343.8	22.0	766	12	BG123960
c 35	342	21.9	711	12	BG646037
c 36	340.8	21.8	361	10	AV557646
c 37	340.4	21.8	796	12	BG645154
c 38	332.6	21.3	1651	11	AV110063
c 39	327.2	21.0	349	10	AV560835
c 40	325.8	20.9	776	12	BG581593
c 41	321.4	20.6	655	14	BQ871920
c 42	318.8	20.4	339	10	AV564807
c 43	317.2	20.3	801	13	BI308524
c 44	316	20.3	324	10	AV560421
c 45	307	19.7	420	10	AV560641

ALIGNMENTS

RESULT 1	AV559121/c	AV559121	585 bp	mrna	linear	EST 07-SEP-2000
LOCUS	AV559121/c	AV559121	Arabidopsis thaliana green silique	Columbia	Arabidopsis	
DEFINITION	thaliana	CDNA clone SQ1110F 3', mRNA sequence.				
ACCESSION	AV559121					
VERSION	AV559121.1	GI:8730547				
KEYWORDS	EST					
SOURCE	Chale cress.					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Asakura, E., Nakamura, Y., Sato, S. and Tabata, S.					
AUTHORS	Epikaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
	1 (bases 1 to 585)					
	Asakura, E., Nakamura, Y., Sato, S. and Tabata, S.					
	A large scale analysis of cDNA in Arabidopsis thaliana: Generation					
	of 12,028 non-redundant expressed sequence tags from normalized and					
	size-selected cDNA libraries					
JOURNAL	DNA Res.	7, 175-180 (2000)				
MEDLINE	20363093					
COMMENT	Contact: Erika Asamizu					
	The First Laboratory for Plant Gene Research					
	Kazusa DNA Research Institute					
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan					
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.					
FEATURES	Location/Qualifiers					
source	1..585					
	/organism="Arabidopsis thaliana"					
	/strain="Columbia"					

```

/db_xref="taxon:3702"
/clone="S011E10F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      160 a 110 c 118 g 197 t
ORIGIN
Query Match      37.5%; Score 585; DB 10; Length 585;
Best Local Similarity 100.0%; Pred. No. 3.9e-102;
Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 925 CATGTATGCGGTTTCGGAACAGAGAAGATGCTTAAAGATTATCTTCTCTTACATTGGA 984
Db 585 CATGTATGCGGTTTCGGAACAGAGAAGATGCTTAAAGATTATCTTCTCTTACATTGGA 526
QY 985 AGAAATCCTCAAAACGATAACTTCACTTTCACGGATCCTTTTCTCACCATTCTTAAT 1044
Db 525 AGAAATCCTCAAAACGATAACTTCACTTTCACGGATCCTTTTCTCACCATTCTTAAT 466
QY 1045 TCTGGCTTGTCAATCGCGGATATCTCTGCTATACCTCCAGAGAAGATTCAAAA 1104
Db 465 TCTGGCTTGTCAATCGCGGATATCTCTGCTATACCTCCAGAGAAGATTCAAAA 406
QY 1105 GCTCCAATGGGATCACTTGAACCAAGAGCTCAGAGAAGATTCGTTGACGAAAGAT 1164
Db 405 GCTCCAATGGGATCACTTGAACCAAGAGCTCAGAGAAGATTCGTTGACGAAAGAT 346
QY 1165 CATAGAAACAAATCGATCAGAGCATTCACAGACATTCCTCGGCTTTCAGTTAAACAAAC 1224
Db 345 CATAGAAACAAATCGATCAGAGCATTCACAGACATTCCTCGGCTTTCAGTTAAACAAAC 286
QY 1225 AATGCTTAAATCTCTTAACTTCCACAAGAAACAGAGACAGCCCTTGTAGACATTGG 1284
Db 285 AATGCTTAAATCTCTTAACTTCCACAAGAAACAGAGACAGCCCTTGTAGACATTGG 226
QY 1285 GATTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTGCGGTGCAACGGTGATTAC 1344
Db 225 GATTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTGCGGTGCAACGGTGATTAC 166
QY 1345 GGATTGAAGTATACAGAGCGCTTGCCAAATATCTGCAATATGGGAGTGGATGTGAAGCA 1404
Db 165 GGATTGAAGTATACAGAGCGCTTGCCAAATATCTGCAATATGGGAGTGGATGTGAAGCA 106
QY 1405 ACTGTTTCAGCCATTGAACAAGCTTCTCGATGTAATGATTGCAAAACATGTGATATT 1464
Db 105 ACTGTTTCAGCCATTGAACAAGCTTCTCGATGTAATGATTGCAAAACATGTGATATT 46
QY 1465 CGACTTTAAATATCAAAAGTTAATTCAATAAACTCGATGATAG 1509
Db 45 CGACTTTAAATATCAAAAGTTAATTCAATAAACTCGATGATAG 1
```

```

RESULT 2
AV557684/c
LOCUS
DEFINITION
Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone S0075c12F 3', mRNA sequence.
ACCESSION
AV557684
VERSION
AV557684.1
KEYWORDS
EST.
SOURCE
thale cress.
```

```

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 606)
Asamizu E., Nakamura Y., Sato S. and Tabata S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
```

```

MEDLINE
COMMENT
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp; URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..606
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0075c12F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      158 a 110 c 124 g 214 t
ORIGIN
Query Match      37.2%; Score 580.2; DB 10; Length 606;
Best Local Similarity 98.7%; Pred. No. 3.2e-101;
Matches 585; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 907 GATGTACCAGAGACTTCTCATGTATGCGGTTTCGGAAACAGAGAGATGCTTTAAAGATTAT 966
Db 606 GATGTACCAGAGACTTCTCATGTATGCGGTTTCGGAAACAGAGAGATGCTTTAAAGATTAT 547
QY 967 CTTTCTCTTACATTTGGAAGAAATCCTGAAACAGATACTTCACTTTACGGAATCCTTT 1026
Db 546 CTTTCTCTTACATTTGGAAGAAATCCTGAAACAGATACTTCACTTTACGGAATCCTTT 487
QY 1027 TCTCACCATTCTCTAAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTCTATACCTC 1086
Db 486 TCTCACCATTCTCTAAATTTCTGGCTTGGTCAATCCGCGGATATTCCTCTCTATACCTC 427
QY 1087 CAGAGAAAGATTCAAAAAGCTCCAATGGATGATCACTTTGAAACAAAGAGCTCAACAAGAA 1146
Db 426 CAGAGAAAGATTCAAAAAGCTCCAATGGATGATCACTTTGAAACAAAGAGCTCAACAAGAA 367
QY 1147 TTGCTTGACGAAAGAAATCATAGAAACAAATTCATGATGAGCATTACAGACATTTCGCGG 1206
Db 366 TTGCTTGACGAAAGAAATCATAGAAACAAATTCATGATGAGCATTACAGACATTTCGCGG 307
QY 1207 CTTTTCAGTTTAAACAACCAATGCTTAAATCTCTTAATCTTAACTTCCACAAGAAACAGAGAC 1266
Db 306 CTTTTCAGTTTAAACAACCAATGCTTAAATCTCTTAACTTCCACAAGAAACAGAGACAG 247
QY 1267 CCTCTTGTAGACGATTGGGATTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTGC 1326
Db 246 CCTCTTGTAGACGATTGGGATTGCTTCAAGACTCTAGTTAATAGCTTCAAGAATCACTGC 187
QY 1327 GGTGCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTGGCCAAATATCTGCAATATG 1386
Db 186 GGTGCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTGGCCAAATATCTGCAATATG 127
QY 1387 GGAGTGATGTGAAGCAAACTGTTTTCAGCCATTGAAACAGCTTGTTCGATGTAATGATTT 1446
Db 126 GGAGTGATGTGAAGCAAACTGTTTTCAGCCATTGAAACAGCTTGTTCGATGTAATGATTT 67
QY 1447 GCAGAAACAAATGTGATATTTCGACTTTTAAATAATATCAAGCTTAAATTTCAATAAAA 1499
Db 66 CCCCAACAAATGTGATATTTCGACTTTTAAATAAAGTTAATTTCAATAAAA 14
```

```

RESULT 3
AV559478/c
LOCUS
DEFINITION
Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone S011B02F 3', mRNA sequence.
ACCESSION
AV559478
VERSION
AV559478.1
KEYWORDS
EST.
SOURCE
thale cress.
```

RESULT 4	AV559545	541 bp	mRNA	linear	EST 07-SEP-2000
LOCUS	AV559545				
DEFINITION	Arabidopsis thaliana green silicles Columbia Arabidopsis thaliana cDNA clone S0119C09F 3', mRNA sequence.				
ACCESSION	AV559545				
VERSION	AV559545.1	GI:8730971			
KEYWORDS	EST.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 541)				
REFERENCE	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.				
AUTHORS	A large scale analysis of cDNA in Arabidopsis thaliana: Generation				
TITLE					

JOURNAL	of 12,028 non-redundant expressed sequence tags from normalized and
MEDLINE	size-selected cDNA libraries
COMMENT	DNA Res. 7, 175-180 (2000)
	20363093
	Contact: Erika Asamizu
	The First Laboratory for Plant Gene Research
	Kazusa DNA Research Institute
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/ .
FEATURES	Location/Qualifiers
source	1..541

/organism="Arabidopsis thaliana"									
/db_xref="taxon:3702"									
/clone="S0119c09f"									
/clone_lib="Arabidopsis thaliana green siliques Columbia"									
/clone_type="green siliques"									
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"									
BASE COUNT	177	a	88	c	128	g	148	t	
ORIGIN									
Query Match	34.7%; Score 541; DB 10; Length 541;								
Best Local Similarity	100.0%; Pred. No. 1e-93;								
Matches 541; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;		
QY	324	GGAAATCCTAGGCC	TGGAGT	TATCAT	TATAAACAGATGGAGAAGATGTTT	TATAAAGG	383		
Db	1	GGAGAATCCTAGGCC	TGGAGT	TATCAT	TATAAACAGATGGAGAAGATGTTT	TATAAAGG	60		
QY	384	AGTTCCTAAGGACTACACT	TAAGAAGCTGTT	TATGTTTCAAAC	TCTTACAATGTGTACT	443			
Db	61	AGTTCCTAAGGACTACACT	TAAGAAGCTGTT	TATGTTTCAAAC	TCTTACAATGTGTACT	120			
QY	444	TGAAATGAAAGTGGCGT	CACAGGAGAAATGCA	AAAGTTGTGAAAGTGGTCT	TAATGA	503			
Db	121	TGGAATGAAAGTGGCGT	CACAGGAGAAATGCA	AAAGTTGTGAAAGTGGTCT	TAATGA	180			
QY	504	TAATATCTTCATCTATT	TGCTGACCATGGAGCT	CTCGCTTAATAGCATGCCCACTGG	563				
Db	181	TAATATCTTCATCTATT	TGCTGACCATGGAGCT	CTCGCTTAATAGCATGCCCACTGG	240				
QY	564	TGATGAAGTTATGGCAAAAGATT	TCAATGAAGTCTTGGAGAAGATGCATAAGAGAAAAA	623					
Db	241	TGATGAAGTTATGGCAAAAGATT	TCAATGAAGTCTTGGAGAAGATGCATAAGAGAAAAA	300					
QY	624	ATACAACAAGATGGTGATCTAT	GTGGAAGCATGTGAATCAGGAAGTATGTTTCAAGGGAT	683					
Db	301	ATACAACAAGATGGTGATCTAT	GTGGAAGCATGTGAATCAGGAAGTATGTTTCAAGGGAT	360					
QY	684	TTTAAAGAAAAATCTCAACATATAC	CGATGCTGCTTAATCTTAAGAGAGACAGCTG	743					
Db	361	TTTAAAGAAAAATCTCAACATATAC	CGATGCTGCTTAATCTTAAGAGAGACAGCTG	420					
QY	744	GGAGTTTACTGTCCTTGAGTCATAT	CTCTCTCTCTCTGAGATTTGGAACTTGCTCGG	803					

OV 744 GGGAGTTTACTGTCTGAGTCATATCCCTCCTCCTCTCTGAGATTGGAACTTGTCTCGG 803

```
Db 421 GGAGTTTACTGTCTGAGTCAATCTCTCTCTCTCTGAGATTGGAACCTTGTCTCGG 480
Qy 804 CGATACATTAGCATCTCTTGGCTTGGAGACAGTGAACCTTCATGACATGAGCAAAAGAC 863
Db 481 CGATACATTAGCATCTCTTGGCTTGGAGACAGTGAACCTTCATGACATGAGCAAAAGAC 540
Qy 864 T 864
Db 541 T 541

RESULT 5
AV561906/c 535 bp mRNA linear EST 07-SEP-2000
LOCUS AV561906 Arabidopsis thaliana green silicles Columbia Arabidopsis
DEFINITION thaliana cDNA clone S016090F 3', mRNA sequence.
ACCESSION AV561906
VERSION AV561906.1 GI:8733332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 535)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp. URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1..535
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S016090F"
/clone_lib="Arabidopsis thaliana green silicles Columbia"
/tissue_type="green silicles"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 149 a 101 c 108 g 177 t
Query Match 34.3%; Score 535; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.5e-92;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1004 ACTTCACCTTCAGGAATCCCTTCCCTCACCACCTCTCTAATCTGGCTTGGTCAATCCGC 1063
Db 535 ACTTCACCTTCAGGAATCCCTTCCCTCACCACCTCTCTAATCTGGCTTGGTCAATCCGC 476
Qy 1064 GCGATATTCCTCTGTATACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGGATCACTTG 1123
Db 475 GCGATATTCCTCTGTATACCTCCAGAGAAAGATTCAAAAGCTCCCAATGGGATCACTTG 416
Qy 1124 AAAGCAAGAAGCTCAGAGAAATTCCTTGACGAAAAGATTCATAGGAAAACAATCGATC 1183
Db 415 AAAGCAAGAAGCTCAGAGAAATTCCTTGACGAAAAGATTCATAGGAAAACAATCGATC 356
Qy 1184 AGAGCATTTACACACATTCCTGGGGCTTTTCAGTTAAACAAACCAATGTCTTAAATCTCTTA 1243
Db 355 AGAGCATTTACACACATTCCTGGGGCTTTTCAGTTAAACAAACCAATGTCTTAAATCTCTTA 296
Qy 1244 TTCCCAAGAACACAGCAGCTCTTGTAGACGATTGGGATTCGCTTCAAGACCTAG 1303
Db 295 CTCCCAAGAACACAGCAGCTCTTGTAGACGATTGGGATTCGCTTCAAGACCTAG 236
```

```
Qy 1304 TTAATAGCTTCAAGAATCACTGCGTGCACCGTGCATTACGGATTGAAGTATACAGGAG 1363
Db 235 TTAATAGCTTCAAGAATCACTGCGTGCACCGTGCATTACGGATTGAAGTATACAGGAG 176
Qy 1364 CGCTTGGCAATATCTGCAATATATGGGAGTGGATGTGAAGCAAACTCTTTTCAGCCATTGAAC 1423
Db 175 CGCTTGGCAATATCTGCAATATATGGGAGTGGATGTGAAGCAAACTCTTTTCAGCCATTGAAC 116
Qy 1424 AGCTTGTCTGATGTAATGATTGCAAAACATGTATTCGACTTTTAAATATCAAA 1483
Db 115 AGCTTGTCTGATGTAATGATTGCAAAACATGTATTCGACTTTTAAATATCAAA 56
Qy 1484 GTTAATTTCAATAAACTCGATAGATGGTGGTTCATGATACTACTTTTAC 1538
Db 55 GTTAATTTCAATAAACTCGATAGATGGTGGTTCATGATACTACTTTTAC 1

RESULT 6
AV562667
LOCUS AV562667 535 bp mRNA linear EST 07-SEP-2000
DEFINITION AV562667 Arabidopsis thaliana green silicles Columbia Arabidopsis
thaliana cDNA clone S0173h03F 3', mRNA sequence.
ACCESSION AV562667
VERSION AV562667.1 GI:8734093
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 535)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp. URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
FEATURES
source
1..535
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0173h03F"
/clone_lib="Arabidopsis thaliana green silicles Columbia"
/tissue_type="green silicles"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 170 a 92 c 114 g 159 t
Query Match 33.9%; Score 528.6; DB 10; Length 535;
Best Local Similarity 99.3%; Pred. NO. 2.4e-91;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 CTCACAAGATTCAGATTCAAGATAGAAAGTTTTCACAACTCTCTAGTCTCTTGGTCCAC 60
Db 1 CTCACAAGATTCAGATTCAAGATAGAAAGTTTTCACAACTCTCTAGTCTCTTGGTCCAC 60
Qy 61 TTTTCAGATTCCTCTTTTCTTCATGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 61 TTTTCAGATTCCTCTTTTCTTCATGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Qy 121 CAATTCGTGAACATATGATGTTGAATCTAGCGACAGAGTGAAGGACACGATGG 180
Db 121 CAATTCGTGAACATATGATGTTGAATCTAGCGACAGAGTGAAGGACACGATGG 180
Qy 181 GCTGTTTTAGTCTGCTGGATCAATGAATATTAATACTACAGGCATCAGCGTGCATATGC 240
Db 181 GCTGTTTTAGTCTGCTGGATCAATGAATATTAATACTACAGGCATCAGCGTGCATATGC 240
```



```

Db 181 GCTGTTTTAGTCTGGATCAAAATGAATATATATAACTACAGGCATCAGGCTGACATATGC 240
Qy 241 CACGCGTATCAGATACTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATG 300
Db 241 CACGCGTATCAGATACTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTATG 300
Qy 301 TATGATGATATCGCGTTTCTCGGAGAAATCCTAGGCGTGGAGTTATCATTAATAACCA 360
Db 301 TATGATGATATCGCGTTTCTCGGAGAAATCCTAGGCGTGGAGTTATCATTAATAACCA 360
Qy 361 GATGGAGAAATGCTTTTATAAGAGAGTTCTTAAGGAGTACACTAAAGAGCTGTTAATGTT 420
Db 361 GATGGAGAAATGCTTTTATAAGAGAGTTCTTAAGGAGTACACTAAAGAGCTGTTAATGTT 420
Qy 421 CAAACTTCTACATGTTGTTTACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATGGCAAA 480
Db 421 CAAACTTCTACATGTTGTTTACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATGGCAAA 480
Qy 481 GTTGTGAAAGTGGTCTTAATGATATATCTTCAATCTATCTATCTGACCATGGAG 535
Db 481 GTTGTGAAAGTGGTCTTAATGATATATCTTCAATCTATCTGACCATGGAG 535

RESULT 7
AV555212 538 bp mRNA linear EST 06-SEP-2000
LOCUS
DEFINITION
thaliana cDNA clone SQ008c10F 3', mRNA sequence.
ACCESSION
AV555212
VERSION
AV555212.1 GI:8726627
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 538)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
source
1. 538
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ008c10F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 168 a 96 c 116 g 158 t
ORIGIN
Query Match 33.6%; Score 524.2; DB 10; Length 538;
Best Local Similarity 98.5%; Pred. No. 1.7e-90;
Matches 529; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 4 ACAAGATCAGATCAAGATAGAGTTTCAACATGCTAGTCTCTGTGTCACCTTT 63
Db 2 ACAAGATCAGATCAAGATAGAGTTTCAACATGCTAGTCTCTGTGTCACCTTT 61
Qy 64 CAGATCTTCTGTTTCTTCATGCTTTGCTTATCTTCTCAGCTGAGTCCCGCAAAACCCAA 123
Db 62 CAGATCTTGGGTGCTCTTTTCTTCTTCTTCTTCTCAGCTGAGTCCCGCAAAACCCAA 121

```

```

Qy 124 TTCTGAAAGATATATGATGTTGAATCTAGCGACAAGAGTGCAAAAGGCACACGATGGCT 183
Db 122 TTCTGAAAGATATATGATGTTGAATCTAGCGACAAGAGTGCAAAAGGCACACGATGGCT 181
Qy 184 GTTTTAGTTGCTCGATCAAAATGAATATATAACTACAGGCATCAGGCTGACATATGCCAC 243
Db 182 GTTTTAGTTGCTCGATCAAAATGAATATATAACTACAGGCATCAGGCTGACATATGCCAC 241
Qy 244 GCCTATCAGATCTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTTATGTAT 303
Db 242 GCCTATCAGATCTCCGAAAGGCGGTTTAAAAGATGAAACATCATTTGTGTTTATGTAT 301
Qy 304 GATGATATCGCGTTTCTCGGAGAAATCCTAGGCGTGGAGTTATCATTAATAACCCAGAT 363
Db 302 GATGATATCGCGTTTCTCGGAGAAATCCTAGGCGTGGAGTTATCATTAATAACCCAGAT 361
Qy 364 GGAGAAGATGTTTATAAGGAGTTCTTAAGGAGTACACTAAAGAGCTGTTAATGTTCAA 423
Db 362 GGAGAAGATGTTTATAAGGAGTTCTTAAGGAGTACACTAAAGAGCTGTTAATGTTCAA 421
Qy 424 AACTTCTACATGTTGTTTACTTGGAAATGAAAGTGGCGTCACAGGAGGAAATGGCAAGTT 483
Db 422 AACTTCTACATGTTGTTTACTTGGAAATGAAAGGCGGTCACAGGAGGAAATGGCAAGTT 481
Qy 484 GTGAAAGTGGTCTTAATGATATATCTTCAATCTATCTGACCATGGAGCTCT 540
Db 482 GTGAAAGTGGTCTTAATGATATATCTTCAATCTATCTGACCATGGAGCTCT 538

RESULT 8
AV557191/c 506 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION
thaliana cDNA clone SQ062f03F 3', mRNA sequence.
ACCESSION
AV557191
VERSION
AV557191.1 GI:8728606
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 506)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
source
1. 506
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ062f03F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 134 a 96 c 105 g 171 t
ORIGIN
Query Match 32.2%; Score 502.8; DB 10; Length 506;
Best Local Similarity 99.6%; Pred. No. 2.1e-86;
Matches 504; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 984 ACAAATCTGAAAGATCAATCTCACTTTCAGGGAATCCTTTCTCACCACATCTCTAA 1043

```

```

Db      506  AAGAAATCCTGAAACGATAAATCTCACTTTCACGGAATCCTTTCTCCTCACCACATCTCTAA 447
Qy      1044 TTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCRAAA 1103
Db      446  TTCTGGCTTGGTCAATCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCRAAA 387
Qy      1104 AGTCCCAATGGGATCTTGAAGCAAGAAAGCTCAGAGAAATGCTTGGACGAAAGAA 1163
Db      386  AGTCCCAATGGGATCTTGAAGCAAGAAAGCTCAGAGAAATGCTTGGACGAAAGAA 327
Qy      1164 TCATAGGAACAATCGATCAGAGCATACAGACATTCCTGGGCTTTTCAGTTAAACAAC 1223
Db      326  TCATAGGAACAATCGATCAGAGCATACAGACATTCCTGGGCTTTTCAGTTAAACAAC 267
Qy      1224 CAATGCTTAAATCTTAACTTCCACAGAACACAGACAGCTCTTGTAGACGATTG 1283
Db      266  CAATGCTTAAATCTTAACTTCCACAGAACACAGACAGCTCTTGTAGACGATTG 207
Qy      1284 GGATGCTTCAAGCTCTAGTTAATAGCTTCAAGAAATCACTGCGTGCACACGCTGCAATTA 1343
Db      206  GGATGCTTCAAGCTCTAGTTAATAGCTTCAAGAAATCACTGCGTGCACACGCTGCAATTA 147
Qy      1344 CGGATGAAGTATACAGAGCGCTTGCATATCTGCAATATGGAGTGGATGTAAGCA 1403
Db      146  CGGATGAAGTATACAGAGCGCTTGCATATCTGCAATATGGAGTGGATGTAAGCA 87
Qy      1404 AACTGTTTCAGCATTTGAACAGCTTGTTCGATGAATGTTGCAAAACAATGTGATAT 1463
Db      86  AACTGTTTCAGCATTTGAACAGCTTGTTCGATGAATGTTGCAAAACAATGTGATAT 27
Qy      1464 TCGACTTTTAAATATCAAGCTTAAT 1489
Db      26  TCGCTTTTAAATATCAAGCTTAAT 1

RESULT 9
AV831492 574 bp mRNA linear EST 01-APR-2002
LOCUS AV831492 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-89-M04 5',
mRNA sequence.
AV831492.1 GI:19873552
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 574)
Seki.M., Narusaka.M., Ishida.J., Kamiya.A., Satou.M., Nakajima.M.,
Oono.Y., Sakurai.T., Carninci.P., Kawai.J., Itoh.M., Ishii.Y.,
Arakawa.T., Shibata.K., Shinagawa.A., Muramatsu.M., Hayashizaki.Y.
and Shinozaki.K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rken.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda phage-1 vector (Garninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
Location/Qualifiers
1..574
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

FEATURES
source

```

```

/clone="RAFL09-89-M04"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/notes="Site_1: BamHI; Site_2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"
BASE COUNT 181 a 102 c 119 g 169 t 3 others
ORIGIN
Query Match
Best Local Similarity 31.5%; Score 492; DB 10; Length 574;
Matches 540; Conservative 97.6%; Pred. No. 2.3e-84;
Mismatches 8; Indels 5; Gaps 4;
Qy      1  CTCACAAAGATCAGATTCAGATAGAGATTTTTCACAAAGATCTCTAGTCTCTTGGTCAC 60
Db      22  CTCACAAAGATCAGATTCAGATAGAGATTTTTCACAAAGATCTCTAGTCTCTTGGTCAC 81
Qy      61  TTTTCAGATCTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120
Db      82  TTTTCAGATCTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 141
Qy      121  CAATTCGTCAGCATATGATGTTGAATCTAGCGACAAAGAGTGCACAAAGAGTGCACAAAGAG 180
Db      142  CAATTCGTCAGCATATGATGTTGAATCTAGCGACAAAGAGTGCACAAAGAGTGCACAAAGAG 201
Qy      181  GCTGTTTATGTTGCTGGATCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
Db      202  GCTGTTTATGTTGCTGGATCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 261
Qy      241  CAGCGTATCAGATATCTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTTTATG 300
Db      262  CAGCGTATCAGATATCTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTTTATG 321
Qy      301  TATGATGATATCGCTTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 360
Db      322  TATGATGATATCGCTTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 381
Qy      361  GATGAGAGAGATGTTTATTAAGAGATCTTAAGAGATCTTAAGAGATCTTAAGAGATCTTA 420
Db      382  GATGAGAGAGATGTTTATTAAGAGATCTTAAGAGATCTTAAGAGATCTTAAGAGATCTTA 441
Qy      421  CAAACTTTTCAATATGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 476
Db      442  CAAACTTTTCAATATGTTTCTCGGAGATCTAGGCTGGAGTTATCATTAATAAACCA 501
Qy      477  CAAAGTTGTGAAAGTCTCTTAATGATATATCTTATGTTGACCATGGA-G 535
Db      502  CAAAGTTGTGAAAGTCTCTTAATGATATATCTTATGTTGACCATGGA-G 561
Qy      536  CTCCTGCTTAAT 548
Db      562  CTCCTGCTTAAT 574

```

```

RESULT 10
AV562308/c 443 bp mRNA linear EST 07-SEP-2000
LOCUS AV562308 Arabidopsis thaliana green silique Columbia Arabidopsis
DEFINITION thaliana cDNA clone S0167H10F 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 443)
Asamizu.E., Nakamura.Y., Sato.S. and Tabata.S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and

```

JOURNAL size-selected cDNA libraries
MEDLINE DNA Res. 7, 175-180 (2000)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..443
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ167H10F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 115 a 90 c 87 g 151 t

ORIGIN

Query Match 28.4%; Score 443; DB 10; Length 443;
Best Local Similarity 100.0%; Pred. No. 6.1e-75;
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 TCTGGCTTGGTCAATCCGCGCGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAA 1104
|||||

Db 443 TCTGGCTTGGTCAATCCGCGCGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAA 384
|||||

QY 1105 GCTCAATGGGATCATTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAAT 1164
|||||

Db 383 GCTCAATGGGATCATTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAAT 324
|||||

QY 1165 CATAGAAACAAATCGATCAGAGCATACAGACATTCGCGGCTTCAGTTAAACAAACC 1224
|||||

Db 323 CATAGAAACAAATCGATCAGAGCATACAGACATTCGCGGCTTCAGTTAAACAAACC 264
|||||

QY 1225 AATGCTTTAATCTCTTAACCTCCACAGAACACACAGGACAGCCCTCTTGTAGACGATTGG 1284
|||||

Db 263 AATGCTTTAATCTCTTAACCTCCACAGAACACACAGGACAGCCCTCTTGTAGACGATTGG 204
|||||

QY 1285 GATTGCTTCAAGACTAGTTAATAGCTTCAAGATCACTCGGTGCAACGGTGCATTAC 1344
|||||

Db 203 GATTGCTTCAAGACTAGTTAATAGCTTCAAGATCACTCGGTGCAACGGTGCATTAC 144
|||||

QY 1345 GGATTGAAGTATACAGGAGCGCTTGCCCAATATCTGCAATATGCGAGTGGATGTGAAGCAA 1404
|||||

Db 143 GGATTGAAGTATACAGGAGCGCTTGCCCAATATCTGCAATATGCGAGTGGATGTGAAGCAA 84
|||||

QY 1405 ACTGTTTCAGCCATTGAACAGCTGTTCGATGTAAATGTTGCAAAACAATGTGATATT 1464
|||||

Db 83 ACTGTTTCAGCCATTGAACAGCTGTTCGATGTAAATGTTGCAAAACAATGTGATATT 24
|||||

QY 1465 CGACTTTAAAAATATCAAAAGTTA 1487
|||||

Db 23 CGACTTTAAAAATATCAAAAGTTA 1

RESULT 11

AV563648/c

LOCUS

DEFINITION AV563648 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana CDNA clone SQ191a03F 3', mRNA sequence.

ACCESSION AV563648

VERSION AV563648.1

KEYWORDS EST

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 437)

REFERENCE

AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..437
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ191a03F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 114 a 86 c 85 g 152 t

ORIGIN

Query Match 28.0%; Score 437; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 8.6e-74;
Matches 437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1059 TCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCCAAAGCTCCAATGGGATC 1118
|||||

Db 437 TCCGCGGATATTCCTCTGCTATACCTCCAGAGAAAGATTCCAAAGCTCCAATGGGATC 378
|||||

QY 1119 ACTTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGGAAACAAT 1178
|||||

Db 377 ACTTGAAGCAAGAAAGCTCAGAGAAATTCCTTGACGAAAGAAATCATAGGAAACAAT 318
|||||

QY 1179 CGATCAGAGCATACAGACATTCCTGCGCTTTCAGTTAAACAACCAATGCTTTAAATCT 1238
|||||

Db 317 CGATCAGAGCATACAGACATTCCTGCGCTTTCAGTTAAACAACCAATGCTTTAAATCT 258
|||||

QY 1239 CTTAACTTCCACAAGAACACAGGACAGCCCTCTTGTAGACGATTGGGATTCCTTCAAGAC 1298
|||||

Db 257 CTTAACTTCCACAAGAACACAGGACAGCCCTCTTGTAGACGATTGGGATTCCTTCAAGAC 198
|||||

QY 1299 TCTAGTTAATAGCTTCAAGAATCACTCGGTGCAACGGTGCATTACGGATTGAAGTATAC 1358
|||||

Db 197 TCTAGTTAATAGCTTCAAGAATCACTCGGTGCAACGGTGCATTACGGATTGAAGTATAC 138
|||||

QY 1359 AGGAGCGCTTCCCAATATCTGCAATATGCGAGTGGATGTGAAGCAAACTGTTTCAGCCAT 1418
|||||

Db 137 AGGAGCGCTTCCCAATATCTGCAATATGCGAGTGGATGTGAAGCAAACTGTTTCAGCCAT 78
|||||

QY 1419 TGAACAAGCTTGTTCGATGTAATGATTGCAAAACAATGTCATATTCGACTTTAAAAATA 1478
|||||

Db 77 TGAACAAGCTTGTTCGATGTAATGATTGCAAAACAATGTCATATTCGACTTTAAAAATA 18
|||||

QY 1479 TCAAAAGTTAATTTCAAT 1495
|||||

Db 17 TCAAAAGTTAATTTCAAT 1

RESULT 12

AV564659

LOCUS

DEFINITION AV564659 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana CDNA clone SQ208b11F 3', mRNA sequence.

ACCESSION AV564659

VERSION AV564659.1

KEYWORDS EST

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 431)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source

1. 431
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ208b11f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 152 a 60 c 104 g 115 t

ORIGIN

Query Match
Best Local Similarity 27.6%; Score 431; DB 10; Length 431;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GGAGATCTAGGCTGGAGTATCAATTAATACCAGATGGAGAGATGTTTATAAGG 383
Db 1 GGAGATCTAGGCTGGAGTATCAATTAATACCAGATGGAGAGATGTTTATAAGG 60

QY 384 AGTCTTAGGACTACATAAGAGCTGTTAATGTTCAAAACCTCTACAATGTGTACT 443
Db 61 AGTCTTAGGACTACATAAGAGCTGTTAATGTTCAAAACCTCTACAATGTGTACT 120

QY 444 TGGAAATGAAGTGGCTACAGAGGAATGCAAGTGTGAAAGTGGTCTTAATGA 503
Db 121 TGGAAATGAAGTGGCTACAGAGGAATGCAAGTGTGAAAGTGGTCTTAATGA 180

QY 504 TAATATCTTATCATTTATCTGACCATGGAGCTCTGGCTTAATAGGATGCCACTGG 563
Db 181 TAATATCTTATCATTTATCTGACCATGGAGCTCTGGCTTAATAGGATGCCACTGG 240

QY 564 TGATGAAGTATGCAAAAGATTTCAATGAAGTGTGGAAGATGCATAAGAGAAAA 623
Db 241 TGATGAAGTATGCAAAAGATTTCAATGAAGTGTGGAAGATGCATAAGAGAAAA 300

QY 624 ATACAACAAGATGGTATCTATCTTGAAGCATGTGAATCAGGAAGTATGTTGAAGGAT 683
Db 301 ATACAACAAGATGGTATCTATCTTGAAGCATGTGAATCAGGAAGTATGTTGAAGGAT 360

QY 684 TTTAAGAAAAATCTCAACATATACGAGTACTCTGCTAATTTCAAGAGACAGCTG 743
Db 361 TTTAAGAAAAATCTCAACATATACGAGTACTCTGCTAATTTCAAGAGACAGCTG 420

QY 744 GGGATTTTACT 754
Db 421 GGGATTTTACT 431

RESULT 13
AV59517/c
LOCUS
DEFINITION
AV59517 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone sq118h05f 3', mRNA sequence.
AV59517
VERSION
AV59517.1 GI:8730943
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 462)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source

1. 462
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ118h05f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 125 a 92 c 86 g 159 t

ORIGIN

Query Match
Best Local Similarity 27.4%; Score 428.2; DB 10; Length 462;
Matches 436; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1097 TTCAAAGCTCCAATGGATCCTTGAAGCAAGAGCTCAGAGCAATTCGCGCTTTCAGTTA 1156
Db 462 TTCAAAGCTCCAATGGATCCTTGAAGCAAGAGCTCAGAGCAATTCGCGCTTTCAGTTA 403

QY 1157 AAAGATCATAGGAACAAATCGATCAGACATTCAGACATTCGCGCTTTCAGTTA 1216
Db 402 AAAGATCATAGGAACAAATCGATCAGACATTCAGACATTCGCGCTTTCAGTTA 343

QY 1217 ACAACACCATGCTCTTAATCTCTTAATCTTCAACAAAGCAAGAGCAAGCTTCTGTAG 1276
Db 342 ACAACACCATGCTCTTAATCTCTTAATCTTCAACAAAGCAAGAGCAAGCTTCTGTAG 283

QY 1277 ACGATTGGGATGCTTCAAGACTCTAGTAACTTCAAGAAATCACTCGGTGCAACGG 1336
Db 282 ACGATTGGGATGCTTCAAGACTCTAGTAACTTCAAGAAATCACTCGGTGCAACGG 223

QY 1337 TGCATTACGGATTCAGATATACAGAGCGCTTGCCTAATATCTGCAATATGGAGTGGATG 1396
Db 222 TGCATTACGGATTCAGATATACAGAGCGCTTGCCTAATATCTGCAATATGGAGTGGATG 163

QY 1397 TGAACAACTGTTTCAGCCATTCAGCCATTCAGCCATTCAGCCATTCAGCCATTCAGCCAT 1456
Db 162 TGAACAACTGTTTCAGCCATTCAGCCATTCAGCCATTCAGCCATTCAGCCATTCAGCCAT 103

QY 1457 GTGATATTCGACTTTAAATATCAAGTAAATTCGAATTAATTCGAATTAATTCGAATTAATTC 1516
Db 102 GTGATATTCGACTTTAAATATCAAGTAAATTCGAATTAATTCGAATTAATTCGAATTAATTC 43

QY 1517 TGGTTCATGATCTACTTTTACATGAAA 1545
Db 42 TGGTTCATGATCTACTTTTACATGAAA 14

RESULT 14
AV810306/c
LOCUS
DEFINITION
AV810306 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-63-G07 3',
mRNA sequence.
AV810306
VERSION
AV810306.1 GI:19844291
KEYWORDS
EST.
SOURCE
thale cress.

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     source           1..431
     /organism="Arabidopsis thaliana"
     /db_xref="taxon:3702"
     /clone="RAFL09-63-G07"
     /clone_lib="RAFL9"
     /dev_stage="plants at various developmental stages from
     germination to mature seeds"
     /lab_host="DH10B"
     /note="Site 1: BamHI; Site 2: SalI; subjected to
     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
     hr) treatments"
BASE COUNT      117 a   89 c   75 g   150 t

Query Match      26.8%; Score 418; DB 10; Length 431;
Best Local Similarity 99.8%; Pred. No. 3.7e-70;
Matches 429; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1110 AATGGGATCACTTGAAGCAAGAGCTCAGAGAAATTCGTCACGAAAGAAATCATAG 1169
      |||||||
Db 431 AATGGGATCACTTGAAGCAAGAGAGCTCAGAGAAATTCGTCACGAAAGAAATCATAG 372
      |||||||

Qy 1170 GAACAATCGATCAGAGCATTCAGACATTCGCGCTTCAGTTAAACAACAATGTT 1229
      |||||||
Db 371 GAACAATCGATCAGAGCATTCAGACATTCGCGCTTCAGTTAAACAACAATGTT 312
      |||||||

Qy 1230 CTTAAATCTCTTAATCTCCACAAGAACAGGACAGCCTCTTGTAGACGATTCGGATTG 1289
      |||||||
Db 311 CTTAAATCTCTTAATCTCCACAAGAACAGGACAGCCTCTTGTAGACGATTCGGATTG 252
      |||||||

Qy 1290 CTTCAAGACTCTAGTTAATAGCTTCAAGATCACTGCGGT-GCAACGGTGCAATTCAGGAT 1348
      |||||||
Db 251 CTTCAAGACTCTAGTTAATAGCTTCAAGATCACTGCGGTGCGCAACGGTGCAATTCAGGAT 192
      |||||||

Qy 1349 TGAAGTATACAGGAGCGCTGCCAATATCTCCATATGGGAGTGGATGTGAGCAAACTG 1408
      |||||||
Db 191 TGAAGTATACAGGAGCGCTGCCAATATCTCCATATGGGAGTGGATGTGAGCAAACTG 132
      |||||||

Qy 1409 TTTCAGGCATTTGAACAAGCTTGTTCGATGTAAATGATTTGCAAAACAATGTGATATTCGAC 1468
      |||||||
Db 131 TTTCAGGCATTTGAACAAGCTTGTTCGATGTAAATGATTTGCAAAACAATGTGATATTCGAC 72
      |||||||

Qy 1469 TTTAAAAATATCAAGTTAAATTTCAATAAAACCTCGATGTAGAGATGGTTGGTTTCATGATA 1528
      |||||||
Db 71 TTTAAAAATATCAAGTTAAATTTCAATAAAACCTCGATGTAGAGATGGTTGGTTTCATGATA 12
      |||||||

Qy 1529 CTACTTTTAC 1538

```

```

Db 11 CTACTTTTAC 2
      |||||||

RESULT 15
AV819966/c
LOCUS AV819966 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-08-D05 3',
DEFINITION mRNA sequence.
ACCESSION AV819966
VERSION AV819966.1 GI:19861922
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 411)
REFERENCE
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     source           1..411
     /organism="Arabidopsis thaliana"
     /db_xref="taxon:3702"
     /clone="RAFL11-08-D05"
     /clone_lib="RAFL11"
     /dev_stage="plants at various developmental stages from
     germination to mature seeds"
     /lab_host="DH10B"
     /note="Site 1: BamHI; Site 2: SalI; subjected to various
     treatments (dehydration, cold, high salt, ABA, heat and UV
     ). Dark-grown plants"
BASE COUNT      105 a   82 c   77 g   147 t

Query Match      26.3%; Score 411; DB 10; Length 411;
Best Local Similarity 100.0%; Pred. No. 8.2e-69;
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1082 ACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCACTTGAAGCAAGAAAGCTCAGA 1141
      |||||||
Db 411 ACCTCCAGAGAAAGATTCAAAAAGCTCCAAATGGGATCACTTGAAGCAAGAAAGCTCAGA 352
      |||||||

Qy 1142 AGAAATTCGTTGACGAAAAGAAATCATAGAAACAAATCGATCAGAGCATTTACAGACATTC 1201
      |||||||
Db 351 AGAAATTCGTTGACGAAAAGAAATCATAGAAACAAATCGATCAGAGCATTTACAGACATTC 292
      |||||||

Qy 1202 TGGCGCTTTTCAGTTAAACAACAACCAATGCTCTTAATCTCTTAACCTCCACAGACAACAG 1261
      |||||||
Db 291 TGGCGCTTTTCAGTTAAACAACAACCAATGCTCTTAATCTCTTAACCTCCACAGACAACAG 232
      |||||||

Qy 1262 GACAGCCTCTTGTAGACGATTTGGGATTCGTTCAAGACTCTAGTTAATAGCTTCAAGATC 1321
      |||||||
Db 231 GACAGCCTCTTGTAGACGATTTGGGATTCGTTCAAGACTCTAGTTAATAGCTTCAAGATC 172
      |||||||

```

Qy 1322 ACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGCA 1381
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
171 ACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACAGGAGCGCTTGCCAAATATCTGCA 112
Qy 1382 ATATGGGAGTGGATGTGAAGCAAACTGTTTCAGGCATTGAACAAAGCTTGTTCGATGTAAT 1441
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
111 ATATGGGAGTGGATGTGAAGCAAACTGTTTCAGGCATTGAACAAAGCTTGTTCGATGTAAT 52
Qy 1442 GATTTCGAAAACAATGTGATATTCGACTTTAAAAATATCAAAAGTTAATTC 1492
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
51 GATTTCGAAAACAATGTGATATTCGACTTTAAAAATATCAAAAGTTAATTC 1

Search completed: May 29, 2003, 17:06:25
Job time : 2094 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 17:06:35 ; Search time 306 Seconds
(without alignments)
3429.514 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVAIEQAQCSM 466

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US0934066/runat_19052003_163659_24347/app_query.fasta_1.647
-DB=N_Geneseq_101002 -QWMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066.CGN.1.1.396.0runat_19052003_163659_24347 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2454	100.0	1401	21	AAC50984
2	2446	99.7	1541	21	Arabidopsis thalia
3	1268	51.7	1742	14	Asparaginylendopep
4	1265	51.5	1749	14	Asparaginylendopep
5	1251	51.0	1323	14	Asparaginylendopep
6	1246	50.8	1485	14	Arabidopsis thalia
7	1245	50.7	1323	14	Asparaginylendopep
8	1220	49.7	1830	14	Asparaginylendopep
9	1215	49.5	1323	14	Asparaginylendopep
10	1080	44.0	1323	14	Asparaginylendopep
11	1080	44.0	1323	14	Asparaginylendopep
12	1080	44.0	1910	14	Asparaginylendopep
13	1019	41.5	1640	14	Asparaginylendopep
14	990	40.3	1152	14	Asparaginylendopep
15	891	36.3	894	14	Asparaginylendopep
16	873.5	35.6	813	14	Asparaginylendopep
17	828	33.7	1936	19	Osteoclast inhibit
18	828	33.7	2030	22	Human FDH02 cDNA.
19	805.5	32.8	753	14	Asparaginylendopep
20	771	31.4	1128	19	Open reading frame
21	771	31.4	1749	19	Human protein comp
22	749	30.5	1453	14	Asparaginylendopep
23	744	30.3	1855	18	Human adrenal glan
24	744	30.3	1855	20	Novel cysteine pro
25	744	30.3	1855	21	Human cysteine pro
26	685	27.9	1336	24	DNA encoding human
c 27	629.5	25.7	2373	23	AA565199
28	613	25.0	1250	14	DNA encoding novel
29	326.5	13.3	981	23	Asparaginylendopep
c 30	300.5	12.2	3172	23	Drosophila melanog
31	269	11.0	1490	21	Arabidopsis thalia
32	254.5	10.4	1491	21	Arabidopsis thalia
33	227.5	9.3	423	16	Enhanced expressio
34	216.5	8.8	640	23	DNA encoding novel
35	195	7.9	2135	22	DNA sequence from
36	178	7.3	220	18	Human cysteine pro
37	178	7.3	220	20	NCP partial cDNA c
38	178	7.3	220	21	Human cysteine pro
39	168	6.8	219	18	Human cysteine pro
40	168	6.8	219	20	Human cysteine pro
41	168	6.8	219	21	NCP partial cDNA c
42	162.5	6.6	461	21	Human secreted pro
43	148	6.0	329	14	Asparaginylendopep
44	146	5.9	15061	22	Human reproductive
45	146	5.9	15061	23	Genomic sequence #

ALIGNMENTS

RESULT 1
AAC50984
ID AAC50984 standard; DNA; 1401 BP.
XX
XX AAC50984;
AC AAC50984;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 66846.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX

PD	06-SEP-2000.
XX	25-FEB-2000; 2000EP-0304139.
PF	
FE	
XX	
PR	25-FEB-1999; 99US-012825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	25-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	04-MAY-1999; 99US-0132407.
PR	05-MAY-1999; 99US-0132484.
PR	06-MAY-1999; 99US-0132485.
PR	08-MAY-1999; 99US-0132486.
PR	07-MAY-1999; 99US-0132487.
PR	11-MAY-1999; 99US-0132486.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUL-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.
PR	01-JUL-1999; 99US-0142154.
PR	02-JUL-1999; 99US-0142055.
PR	06-JUL-1999;
PR	08-JUL-1999;
PR	09-JUL-1999;
PR	12-JUL-1999;
PR	13-JUL-1999;
PR	14-JUL-1999;
PR	15-JUL-1999;
PR	16-JUL-1999;
PR	16-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	19-JUL-1999;
PR	20-JUL-1999;
PR	20-JUL-1999;
PR	20-JUL-1999;
PR	20-JUL-1999;
PR	21-JUL-1999;
PR	21-JUL-1999;
PR	21-JUL-1999;
PR	21-JUL-1999;
PR	21-JUL-1999;
PR	22-JUL-1999;
PR	22-JUL-1999;
PR	22-JUL-1999;
PR	22-JUL-1999;
PR	22-JUL-1999;
PR	23-JUL-1999;
PR	23-JUL-1999;
PR	23-JUL-1999;
PR	26-JUL-1999;
PR	27-JUL-1999;
PR	27-JUL-1999;
PR	27-JUL-1999;
PR	27-JUL-1999;
PR	28-JUL-1999;
PR	28-AUG-1999;
PR	28-AUG-1999;
PR	02-AUG-1999;
PR	02-AUG-1999;
PR	03-AUG-1999;
PR	04-AUG-1999;
PR	04-AUG-1999;
PR	05-AUG-1999;
PR	05-AUG-1999;
PR	06-AUG-1999;
PR	06-AUG-1999;
PR	09-AUG-1999;
PR	09-AUG-1999;
PR	10-AUG-1999;
PR	10-AUG-1999;
PR	11-AUG-1999;
PR	12-AUG-1999;
PR	13-AUG-1999;
PR	13-AUG-1999;
PR	16-AUG-1999;
PR	17-AUG-1999;


```

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 2,27e-231 Length: 1401
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-934-066-2 (1-466) x AAC50984 (1-1401)

Qy 1 MetSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 1 ATGTCTAGTCTCTTGGTCACTTTTCAGATTCTTTTCTTCTCATGCTTTTCTTATCTTC 60

Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
Db 61 TCAGCTGAGTCCGCCAAACCCAAATTCCTGAACGATAATGATGTGAATCTAGCGCAAG 120

Qy 41 SerAlaLysGlyThrArgTTPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 121 AGTCAAAGGCACACACATGGGCTGTTTACTGCTGATCAATGATGATATATACTAC 180

Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 181 AGGCATCAGGCTGACATATGCCACGCTATCAGATACCTCCGAAAGCGGTTTAAAGAT 240

Qy 81 GluAsnIlelleValPheMetTyrAspIleAlaPheSerSerGluAsnProArgPro 100
Db 241 GAAACATCATGTGTTTATGTATGATGATATCCGTTTCTCCGAGAAATCTTAGGCCT 300

Qy 101 GlyValIlelleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 301 GGAGTTATCATTAATAACCGATGGAGAGATGTTTATAAGGAGTTCCTTAAGGACTAC 360

```

RESULT 2

```

Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db ACTAAAGAAGCTGTTAATGTTCAAACTTCTACAATGTGTACTTGGAAATGAAGTGGC 420

Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db GTACAGGAGGAATGGCAAGTGTGAAAGTGGTCTTAATGATAATATATCTTCATCTAT 480

Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db TATGCTGACCATGGAGCTCTGCTTAATAGCATGCCACTGGTATGAAGTATATGCA 540

Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db AAAGATTCAATGAAGTCTTGGAGAGATGCATAAGAGAAAAAATACAAAGATGCTG 600

Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db ATCTATGTTGAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 660

Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysPro 240
Db AACATATACCGAGTACTGCTGCTAATTTCTAAAGAGAGCAGCTGGGAGTCTTACTGCTC 720

Qy 241 GluSerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db GAGTCATATCTCTCTCTCTCTGAGATTGGAACCTTCTCGGCGATACATTTAGCATC 780

Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db TCTTGGCTTGAGGACAGTGCCTTCATGACATGAGCAAGAGACTTTTGGAGCAACAATAC 840

Qy 281 HisValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db CACGTGCTAAAGAGACAGTAGGATCTGATGTACCAGACTTCTCATGTATGCGGTTTC 900

Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db GGAACAGAGAAGATGCTTAAAGATTATCTTCTCTTACATTTGGAAGAAATCTCTGAAAC 960

Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db GATAACTTCACTTTTCACGGAATCTTCTTCTCAACAATCTCTAATCTTGGCTTGGTCAAT 1020

Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db CCGCGCGATATTCCTCTGCTATACCTCCAGAGAAGATTCAAAAAGCTCCCAATCGGATCA 1080

Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db CTTGAAAGCAAGAAAGCTCAGAGAAATGCTTGACCAAGAAATCATAGGAAACAATC 1140

Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db GATCAGAGCATTCACAGACATTCGCGGCTTTCAGTTAAACAAACCAATGTCTTAAATCTC 1200

Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCysPheLysThr 420
Db TTAACCTCCACAGACACACAGACACGCTTCTGTAGACCATTTGGGATGCTCTCAAGACT 1260

Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db CTAGTTAATAGTTCGAAGATCACTGCGGTGCAACGGTGCATTTACGCGATTTGAAGTATACA 1320

Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGTAAGCAAACTGTTTACGCCATT 1380

Qy 461 GluGlnAlaCysSerMet 466
Db 1381 GAACAAGCTTGTTCGATG 1398

```

AAC35910
ID AAC35910 standard; DNA: 1541 BP.
XX AC AAC35910;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 11859.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

```

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154038.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Alignment Scores:

```

Pred. No.: 1.58e-230
Score: 2446.00
Percent Similarity: 99.79%
Best Local Similarity: 99.79%
Query Match: 21
DB: 21

```

```

Length: 1541
Matches: 465
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0

```

US-09-934-066-2 (1-466) x AAC35910 (1-1541)

```

Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 59 ATGCTAGTCTCTGTGCTACCTTCAGATTCTGTTTCTTCATGCTTTGCTTATCTC 118
Qy 21 SerAlaGlySerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
Db 119 TCAGCTAGTCCCGCAAAACCAATTCCTGAACGATTAATGATGTTGAATAGCGACAAG 178

```

```

Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTrpTyrAsnTyr 60
Db 179 AGTCAAAAGGCACACGATGGCTGTTTGTGTCGATCAATGAATATATACTAC 238
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 239 AGGCATCAGGTCACATATCCACGCGTATCAGATACCTCCGAAAGCGGTTTAAAGAT 298
Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 299 GAAACATCATTTGTTTATGATGATGATATCCGCTTTTCTCGGAGATCCTAGGCGCT 358
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 359 GGAGTTATCATTAATAAACAGATGGAGAGATGTTTATAAAGAGTTCCTTAAGGACTAC 418
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 419 ACTAAAGAGCTTAAATGTTCAAAACTTCTACAACTGTTACTTGGAAATGAAAGTGC 478
Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 479 GTCACAGAGAAATGCAAAAGTTGCAAAAGTGTCTTAATGATATATATCTCATCTAT 538
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 539 TATGCTCACCATGAGCTCTGGCTTAATAGCGATGCCACTGCTGATGAAGTTATGGCA 598
Qy 181 LysAspPheAsnGluValLeuGluLysMethHisLysArgLysLysTyrAsnLysMetVal 200
Db 599 AAAGATTCAATGAAGTCTTGGAGAGATGCATAAGAGAGAAATATACACAAGATGGTG 658
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 659 ATCTATGTTGAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAGAAAAATCTC 718
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db 719 AACATATACGACGTGACTGCTGCTTAATCTAAAGAGAGACGCTGGGAGTTTACTGCTCT 778
Qy 241 GluSerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db 779 GAGTCATATCTCTCTCTCTCTCTGAGATTGGAATTTGCTCGCGCATACATTTAGCATC 838
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyr 280
Db 839 TCTTGGCTTGAGGACAGTACCTTCATGACATGACCAAGAGACTTTGGAGCAACAATAC 898
Qy 281 HisValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 899 CACGTTGTAAGAGAGAGTAGGATCTGATGTACCAGAGACTTCTCATGTATGCGGTTTC 958
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 959 GGAACAGAGAGATGCTTAAAGATTTATCTTCTCTTACATTTGGAAGAAATCTCTGAAAC 1018
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 1019 GATAACTTCACTTTCACGGAATCCTTTTCTCACCATCTCTAATCTCGCTTGGCTCAAT 1078
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1079 CCGCGCATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAAGCTCCAAATGGGATCA 1138
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1139 CTTGAAAGAGAGAGAGCTCAGAGAAATGCTTGCAGAAAGAAATCATAGGAAACAATC 1198
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1199 GATCAGAGCATTACAGACATTCGCGGCTTTCAGTTAAACAAACCAATGCTTAAATCTC 1258

```

Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
 Db 1259 TTAACCTTCACAGAACACAGGACAGCCTCTGTAGACGATTGGGATTGCTTCAAGACT 1318
 Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
 Db 1319 CTAGTTAATAGCTTCAGAAATCAGTCGCGTCAACGGTGCATTACGGATTGAAGTATACA 1378
 Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
 Db 1379 GGAGCGCTTGCCTCAATATCTGAATATGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1438
 Qy 461 GluGlnAlaCysSerMet 466
 Db 1439 GAACAGCTTGTCTCGATG 1456

RESULT 3

AAQ50570
 ID: AAQ50570 standard; cDNA to mRNA; 1742 BP.

AC AAQ50570;

DT 24-MAY-1994 (first entry)

XX Asparaginyl endopeptidase clone 101.

XX Asparaginyl endopeptidase; Canavalia ensiformis; seed;

KW L-asparagine; primer; PCR; protein fragmentation;

KW peptide synthesis; ss.

XX Canavalia ensiformis.

XX Key Location/Qualifiers
 FT CDS 207..1529
 FT /*tag a

XX JP05276960-A.

XX 26-OCT-1993.

XX 07-AUG-1992; 92JP-0231602.

XX 07-FEB-1992; 92JP-0056023.

XX (SYUZ/) SYUZO T.

XX WPI; 1993-373587/47.

XX P-PSDB; AAR43035.

XX New gene for encoding asparaginyl endo-peptidase - comprises 8

XX specified DNA sequences

XX Disclosure; Page 17-19; 35pp; Japanese.

XX A gene encoding asparaginyl endopeptidase is claimed.

XX 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

XX derived from a seed of Canavalia ensiformis which selectively

XX hydrolyses C-terminus amide bond of L-asparagine residue

XX (see AAR43033 and AAR43041).

XX The enzyme is useful for protein fragmentation and enzymatic

XX peptide synthesis.

XX The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used

XX in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,

XX by PCR.

XX Sequence 1742 BP; 507 A; 338 C; 395 G; 502 T; 0 other;

Alignment Scores:

Pred. No.: 1,068-114 Length: 1742

Score: 1268.00 Matches: 243

Percent Similarity: 65.97% Conservative: 71

Best Local Similarity: 51.05% Mismatches: 140

Query Match: 51.67% Indels: 22

DB: 14 Gaps: 4
 US-09-934-066-2 (1-466) x AAQ50570 (1-1742)
 Qy 2 SerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuLephe--- 20
 Db 75 TCAATTACCATGAACCCCTTTTCGATCCTTTTCCTCCGACCTCCCAAGCTTCCAGATCTTCCAAAGCA 134
 Qy 21 -----SerAlaIleSerArgLysThrGlnLeu 29
 Db 135 TCCGGTGACCATGACGATGTTCCCGATTGCCCTCCGAAAGCTTCCAGATCTTCCAAAGCA 194
 Qy 30 LeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrpAlaVal 49
 Db 195 CCCGAGGAAACAGAC-----GAGGGGACAGGTGGCCGTT 230
 Qy 50 LeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAla 69
 Db 231 TTAATTGCTGGTTCCCAATGGTTACTGGAATTACAGGCATCAGTCTGATGTATGCCATCGG 290
 Qy 70 TyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleValPheMetTyrAsp 89
 Db 291 TATCAAGTGTGAGGAAAGGTGCTCTGAAAGAGAAATATTGTTGTTTATGTATGAT 350
 Qy 90 AspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAspGly 109
 Db 351 GACATTGCTTTCCATAAAGAAACCCACGCGCTGGAATCATCTTAACAGTCCACATGGA 410
 Qy 110 GluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsn 129
 Db 411 GATGATGTTTACAAGGAGTACCTTAAGGATTACGTTGGTAAAGATGTTACTGTTAACAAAC 470
 Qy 130 PheTyrAsnValLeuLeuGlyAsnGluSerGlyValIleThrGlyLysAsnGlyLysValVal 149
 Db 471 TTTTGTAGCTGCTACTTGGAAATAAGTCAAGCTATTACCGCGCGTGTAGTGGAGGTGTC 530
 Qy 150 LysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeu 169
 Db 531 AATAGTGTCCCAATGATCATATATTATTTACTCTGATCATGGGCGTCCGGAGTG 590
 Qy 170 IleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGlyLys 189
 Db 591 CTGGGATGCCCTACTAGTCCGTTCTGTATGATCTGATCTGATGAACTTTGAAGAG 650
 Qy 190 MethLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGly 209
 Db 651 AAGCATGCTTCTGGAACATATAAAGCCTTGTATTTTATCTAGAGGCATGTGAATCTGGG 710
 Qy 210 SerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsn 229
 Db 711 AGTATCTTTGAAGCTCTTCTCTGGAAGGTCTAAATATCTATGCAACAACACAGCTCAAAAT 770
 Qy 230 SerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProProSerGlu 249
 Db 771 GCAAAAGAAACGACTTGGGAAACATATTTCCTCGGGAGTCTCCTAGTCTCTCCCCACAA 830
 Qy 250 IleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHis 269
 Db 831 TATGAAACCTTCCCTGGGTGACCTGTACAGTGTGCTTGGATGGAAGACAGTGCATACAC 890
 Qy 270 AspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLys---ArgArgValGly 288
 Db 891 AATTATCAACAGAAACCTCTACATCAACATACGAATCGGTCAAGCAAGGACTATCAAT 950
 Qy 289 SerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAsp 308
 Db 951 GCAAAATTCATTTTATGTTCTCAGTGATGCGATGTTGGTACATAGAGCTTAGCAAAAT 1010
 Qy 309 TyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSer 328
 Db 1011 AATCTCTTCCTATATTGGGTGCAAAATCTCTGCAAAATGATAATTTTACTTTTCTGGATAAA 1070
 Qy 329 PheSerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyr 348

Db 964 AGGACCATAGTGGAGTTCTACTATGCTCTCAGCTGATGCGAGTATGCTGATATAGCG 1023
 Qy 305 MetLeuLysAspTyrLeuSerSerTyrTleGlyArgAsnProGluAsnAspAsnPheThr 324
 Db 1024 CACGGATGATGTTCTCTCTCTATATGGGTACAATCTGCTAATGATAATTTTACC 1083
 Qy 325 PheThrGlu-----SerPheSerSerProIleSerAsnSerGlyLeuValAsnProArg 342
 Db 1084 TTTCGATGATAAAGCTCTCTGAGGTACCT-----TCAAAAGCAGTCAACCAACG 1134
 Qy 343 AspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGlu 362
 Db 1135 GATGCTGATCTCATCTTCCTGGGAGAAGTTCGCAAGCTCTGAGGTTCTCCGAG 1194
 Qy 363 SerLysClnAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGln 382
 Db 1195 AAAATGCTGCTCAGAAACAGTTCTGGAAGTAATGCTCACAGGATGCATATAGACAAC 1254
 Qy 383 SerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuThr 402
 Db 1255 GGTGTGGAACCTTATTGGGAGCTTTTATTGGCATTTGAAAGGTCCAAAGTACTGGAT 1314
 Qy 403 SerThrArgThrThrClyClnProLeuValAspTrpAspTrpAspCysPheLysThrLeuVal 422
 Db 1315 GCCTTAGACCGCTGGATGCGCACTTGTGACTGGGACTGCTGCAAAACCATGGTA 1374
 Qy 423 AsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAla 442
 Db 1375 AGGACATTTGAGACATATTTGGATCCTTGTCTCAGTATGGATGGAACATATGAGGTCC 1434
 Qy 443 LeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGln 462
 Db 1435 TTTCGCAACATCTGCAACGAGGAATTAAGAATGACCAAAATGGCCGATGCTCAGCACAA 1494
 Qy 463 AlaCys 464
 Db 1495 GCTTGT 1500
 RESULT 5
 AAQ50562
 ID AAQ50562 standard; cDNA to mRNA; 1323 BP.
 AC AAQ50562;
 DT 24-MAY-1994 (first entry)
 DE Asparaginylendopeptidase sequence (4).
 KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis; ss.
 OS Canavalia ensiformis.
 PN JP05276960-A.
 PD 26-OCT-1993.
 PF 07-AUG-1992; 92JP-0231602.
 PR 07-FEB-1992; 92JP-0056023.
 PA (SYUZ/) SYUZO T.
 DR WPI; 1993-373587/47.
 PT New gene for encoding asparaginyl endo-peptidase - comprises 8
 PT specified DNA sequences
 PS Claim 1; Page 13-14; 35pp; Japanese.
 CC A gene encoding asparaginylendopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

CC derived from a seed of Canavalia ensiformis which selectively
 CC hydrolyses C-terminus amide bond of L-asparagine residue
 CC (see AAR43033 and AAR43041).
 CC The enzyme is useful for protein fragmentation and enzymatic
 CC peptide synthesis.
 CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
 CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
 CC by PCR.
 XX
 SQ Sequence 1323 BP; 375 A; 255 C; 328 G; 365 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,43e-113 Length: 1323
 Score: 1251.00 Matches: 233
 Percent Similarity: 70.05% Conservative: 64
 Best Local Similarity: 54.95% Mismatches: 121
 Query Match: 50.98% Indels: 6
 DB: 14 Gaps: 3
 US-09-934-066-2 (1-466) x AAQ50562 (1-1323)
 Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
 Db 7 GGTACCAAGGTGGGCACTCTTCGCGCGTTCCTCAATGGCTACTGGAATACAGGCATCAG 66
 Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
 Db 67 GCTGATATTTGTCATGCGTATCAACACACTCAGGAAGGTGGCTGAAAGAAATAATATT 126
 Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
 Db 127 ATTTCTTTTATGATGATGACATTTGCTTTCATGAGGAGAACCCAGACCTGGGTCTATC 186
 Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
 Db 187 ATTAACAACACAGATGGGACTGATGTTTATGAAGGAGTTTCCAAAGGATTACACTGGCCAA 246
 Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
 Db 247 GATCTTACTGCTGNAACCTTCTACGCTGTTTACTTGGAAATAAGTCGCACCTTAGCGGT 306
 Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
 Db 307 GGCATGGGAGGTTTGAACAGTGGCCCGCATGCTATATTGTATTCTATAGTCTAT 366
 Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
 Db 367 CATGGAGTCCAGAGTTCGCGGATGCTTCTTACTTATATCCATCTCATCTCTG 426
 Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
 Db 427 AATGATCTTGAAGAAAAAACAATGCTCCGGAACATATATAAAGCCTAGTATTTATCTT 486
 Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
 Db 487 GAGCATGTGAATCTGGAGTATCTTTTGAAGGCTTCTTCTCTGAAGATCTCAATATTTAT 546
 Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
 Db 547 GCAACAACAGCTTCGAATGCAAGAAAGACAGCTGGGGAACATATTTGCCAGGGAGTAT 606
 Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
 Db 607 CCCAGTCTCCCCCAGANATTTCACTTCTTGGGAGACCTGTACAGTGTCTTGGATG 666
 Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlyThrHisValVal 283
 Db 667 GAACAGATGCACAGACACAATTTGCCAACTGAACTTTGAACCAACAATATAAATTTGGT 726
 Qy 284 LysArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
 Db 727 AAGAGAGGAGCAATTAGTGAGGTTTCATCTACTATGCTCTCAGCTGATGCAGTATGCTGAT 786

```
QY 303 GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn 322
Db 787 ATAGGCTCAGCGATGCTTCTTCTCCTATATTTGGGTACAATCCTGCTAATGATAAT 846
QY 323 PheThrPheThrGlu-----SerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 847 TTTACCTTTGCGATGAAACTCCTTGAGCTACCT-----TCAAAAGCAGTCAAC 897
QY 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 898 CAACGGGATCCTCATCTCATCCATTTCTGGGAGAAGTTGCGCAAGCTCCTGAGGGTTCT 957
QY 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 958 CCCGAAAAAATCCTCTCAGAACCAAGTTGTGGAAGTAATGCTCACAGGATGCATATA 1017
QY 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1018 GACAACGGTGTGAACCTTATTGGAGCTTTATTGGCATTGAAAGGGTCCAAAAGTA 1077
QY 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1078 CTGGATGCTGTAGACGGCTGGAATGGCATTGTTGATCAGTGGGACTGCTCGTGAACACC 1137
QY 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1138 ATGCTAAGGACATTTGAGACATATTCTGGACATATTCTGCTCAGTATGGGATGAACATATG 1197
QY 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1198 AGTCTCTTCCAACATCTCCAACGAGGAATTAAGAAATGACCAATGGCCGATGCCTCA 1257
QY 461 GluGlnAlaCys 464
Db 1258 GCACAAGCTTGT 1269

RESULT 6
AAC42934
ID AAC42934 standard; DNA; 1485 BP.
XX AC
AC AAC42934;
XX XX
DT 17-OCT-2000 (first entry)
XX XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37409.
XX XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX XX
OS Arabidopsis thaliana.
XX XX
EP1033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PP 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
```



```

Db 664 ATGGAAGACAGTGGACATACAACTTACGACAGAACTTACATCAACAATAGCAATCG 723
Qy 283 ValLys---ArgArgValGlySerValProGluThrSerHisValCysArgPheGly 301
Db 724 GTCAAGACAGGACTCAATCAATGAAATCAATTTATGTTGTTCTCAGGTATGCAATGGT 783
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321
Db 784 GACATAGAGCTTAGCAAAAATAATCTCTCTATATTTGGGTACAAATCTCGCAAAATGAT 843
Qy 322 AsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnPro 341
Db 844 AATTTTACTTTTGTGGATAAA---AACTCGTTGTCGCCACCTTCAAAAAGCAGTAACCAA 900
Qy 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 901 CGTGATGCTGATCTCGTCCATTCTGGGATAAGTTCGCCAAAGCTCCTGAGGGTCTGCT 960
Qy 362 GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGluIleAsp 381
Db 961 AGCAAGCTGCACGACAGGAACAAGTTCTGGAAGCAATGCTCACAGATCATATAGAT 1020
Qy 382 GlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeu 401
Db 1021 GACAGCATGAACCTTATGGAAGCTCTTATTTGGCATGAAAGGGCCCAAGAGTGCTC 1080
Qy 402 ThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrPaspCysPheLysThrLeu 421
Db 1081 AGCAGTGTAGACCTCTCGGCAAGCACTGTGTGATGATGGGACTGCTTTAAACACATG 1140
Qy 422 ValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGly 441
Db 1141 GTTAGACATTTTGAGACACATTCGCGATCCTGCTCAGTATGGGATGAAGCATATGAGG 1200
Qy 442 AlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGlu 461
Db 1201 TCGTTGCAAACTCTGCAACGCTGTATTCAGAAGGAGCAATGGCTGAGGCTCAGCA 1260
Qy 462 GlnAlaCys 464
Db 1261 CAAGCGTGC 1269

RESULT 8
ID AAQ50572
AC AAQ50572;
DT 24-MAY-1994 (first entry)
DE Asparaginylendopeptidase clone 103.
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
XW L-asparagine; primer; PCR; protein fragmentation;
XX peptide synthesis; ss.
OS Canavalia ensiformis.
Key Location/Qualifiers
CDS 86..1408
/*tag= a
JP05276960-A.
26-OCT-1993.
07-AUG-1992; 92JP-0231602.
07-FEB-1992; 92JP-0056023.
(SYUZ/) SYUZO T.

```

```

DR WPI; 1993-373587/47.
XX P-PSDB; AAR43037.
PT New gene for encoding asparaginyl endo-peptidase - comprises 8
PS specified DNA sequences
XX Disclosure; Page 20-22; 35pp; Japanese.
XX A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX SQ Sequence 1830 BP; 540 A; 348 C; 411 G; 531 T; 0 other;

Alignment Scores:
Pred. No.: 5,92e-110 Length: 1830
Score: 1220.00 Matches: 227
Percent Similarity: 68.37% Conservative: 67
Best Local Similarity: 52.79% Mismatches: 130
Query Match: 49.71% Indels: 6
DB: 14 Gaps: 3

US-09-934-066-2 (1-466) x AAQ50572 (1-1830)
Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuAlaGlySerAsnGluTyr 57
Db 74 ACCGATCAACAACATTCGCGGTACCAAGATGGCCATCTTAATCGCGCGCTCCAATGGCTAC 133
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77
Db 134 TGGAAATTACAGGCATCAGGCTGATATTGTCATGCTATCAATCAATACAGGAAAGTGGC 193
Qy 78 LeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn 97
Db 194 CTGAAGAAGAAATATATTGTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 253
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 254 CCAAACCTGGAGTCAATCAACAAACACAGATGGGATGATGATGATGATGATGATGATGATGAT 313
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 314 AAGGATTAACCTGGCAGCATGTTACTCTCTATCAATCTCTATCTCTATCTCTATCTCTATCT 373
Qy 138 GluSerGlyValThrGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIle 157
Db 374 AAATCAGCATTTACAGGTGGCAGTGGGAAAGTGTGTAACACAGTGGCCCTGATGATGATGAT 433
Qy 158 PheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 434 TTCAATATTTATAGTATGATCATGCGAGTCCAGAGTCTCTGGGATGCGCTGCTGCTGCTTTC 493
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsn 197
Db 494 TTATATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 553
Qy 198 LysMetValIleTyrValIleAlaCysGluSerGlySerMetPheGlyIleLeuLys 217
Db 554 AGCCTTGATTTTATCTTGGAGCATGTCAATCTGGAGTATCTTTGAAGGCTTCTTCTCTCT 613
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 614 GAAGATATCAATATTTATGCAACAACAGCTTCCAAATGCAGAAAGAACAGCAGCTGGGAACA 673
Qy 238 TyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThr 257

```

```

Db 674 TATTGCCCGAGGAGGATCCCTCCCTCCCGAGAAATATTCAACTTGGTGGTGACGAG 733
Qy 258 PheSerIleSerThrLeuGluAspSerLeuHisAspMetSerLysGluThrLeuGlu 277
   ::::::::::::::::::::::::::::::::::::
Db 734 TACAGTGTGCTGGATGAGACACATGACAGGAAGAAATTTGGACAGAACTTTGCAC 793
Qy 278 GlnGlnThrHisValValLysArgArg---ValGlySerAspValProGluThrSerHis 296
   ::::::::::::::::::::
Db 794 CAACAATATGAATTTGGTTAAAGAGAGAGACTATTAAACGGAAGTATATACCATGCTCTCAC 853
Qy 297 ValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArg 316
   ::::::::::::::::::::
Db 854 GTGATCGCATGTGGTGAATTAAGTCTACAGGATGATGTTCTCTCTCTATATTTGGGTACA 913
Qy 317 AsnProGluAsnAspAsnPheThrPheThrGlu-----SerPheSerSerProIleSer 334
   ::::::::::::::::::::::::::::
Db 914 AATCCTGCTAATGATAATTTTACCTATGTGGTGAAGTCTCTCTGAGGTCACT----- 967
Qy 335 AsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGln 354
   ::::::::::::::::::::
Db 968 ---TCAAAAGCAATCAGCCAACTGTTGCTGATCTCATCCATTTTGGGAGAAAGTTCCGC 1024
Qy 355 LysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLys 374
   ::::::::::::::::::::
Db 1025 AAGCTCCTGAGGGTTCTACAGGAAGAGATGCTGCTCAGAAACAATTTCTGGAAGTAAG 1084
Qy 375 AsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGln 394
   ::::::::::::::::::::
Db 1085 TCTCACAGATGCATATAGACAACACTGTGCTGATCTCATCCATTTTGGGAGTCTTTTATTTGCCATT 1144
Qy 395 ThrAsnValLeuAsnLeuThrSerThrArgThrGlyGlnProLeuValAspAsp 414
   ::::::::::::::::::::
Db 1145 GAAAGGGTCCAGAGTACTCAATGCTGTAGACCGCTGGAATGGCACTTGTGTATGAC 1204
Qy 415 TrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHis 434
   ::::::::::::::::::::
Db 1205 TGGGACTGCTGAAAATATGTTAAGGACTTTTGAGACATATTGTGGATCTTGTCTCAG 1264
Qy 435 TyrGlyLeuLysThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLys 454
   ::::::::::::::::::::
Db 1265 TATGGGATGAATATATATAGGTCCTTTGCAACATCTGCAATGCAAGAAATTAAGATGAC 1324
Qy 455 GlnThrValSerAlaIleGluGlnAlaCys 464
   ::::::::::::::::::::
Db 1325 CAAATGGCTGATGCCTCAGCACAAAGCTTGT 1354

RESULT 9
ID AAQ50561 standard; cDNA to mRNA; 1323 BP.
XX
AC AAQ50561;
XX
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase: sequence (3).
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
XX L-asparagine; primer; PCR; protein fragmentation;
XX peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
PN JP05276960-A.
XX
PD 26-OCT-1993.
XX
PF 07-AUG-1992; 92JP-0231602.
XX
PR 07-FEB-1992; 92JP-0056023.
XX (SYUZ/) SYUZO T.
XX
XX WPI; 1993-373587/47.

```

```

XX New gene for encoding asparaginyl endo-peptidase - comprises 8
PT specified DNA sequences
XX
PS Claim 1; Page 13; 35pp; Japanese.
XX
CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX
SQ Sequence 1323 BP; 390 A; 244 C; 312 G; 377 T; 0 other;

Alignment Scores:
Pred. No.: 1-18e-109 Length: 1323
Score: 1215.00 Matches: 226
Percent Similarity: 68.63% Conservative: 65
Best Local Similarity: 53.30% Mismatches: 127
Query Match: 49.51% Indels: 6
DB: 14 Gaps: 3

US-09-934-066-2 (1-466) x AAQ50561 (1-1323)
Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 7 GGTACAGATGGGCCATCTTAATCGCGGCTCCCAATGGTACTGGAAATTACAGGCATCAG 66
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
   ::::::::::::::::::::::::::::::::::
Db 67 GCTGATATTGTCATCGGTATCAAAATACATGAGGAAGGTGGCCCTGAAAGAAATAATT 126
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
   ::::::::::::::::::::::::::::::::::
Db 127 ATGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATC 186
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
   ::::::::::::::::::::::::::::::::::
Db 187 ATTAACAACACAGATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 246
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
   ::::::::::::::::::::::::::::::::::
Db 247 GATGTTACTGCTCATAACTTCTATGCTGTTATATCTTGGAAATAAATCAGCAGTACAGGT 306
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
   ::::::::::::::::::::::::::::::::::
Db 307 GGCAGTGGGAAGGTTGTAACAGTGGCCCTGATGATGATGATGATGATGATGATGATGATGAT 366
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
   ::::::::::::::::::::::::::::::::::
Db 367 CATGGAGGTCCAGAGGTTCTTGGGATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 426
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
   ::::::::::::::::::::::::::::::::::
Db 427 ATTGAAGTCTTGAAGAAAAACATGCTTCTGGAACCGTATATAAAGCCCTTGTATTTATCTT 486
Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
   ::::::::::::::::::::::::::::::::::
Db 487 GAGGCATGTGAATCTGGAGTACTTTTGAAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 546
Qy 224 AlaValThrAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
   ::::::::::::::::::::::::::::::::::
Db 547 GCAACAACACTTCCCAATGCAGAAAGAACAGCAGTGGGAAACATATTTCCCGCAGCGAGGAT 606
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
   ::::::::::::::::::::::::::::::::::
Db 607 CCCAGTCTCCCGCCAGCAATATTCACACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 666

```

```

QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
DB 667 GAAGACAGTGCAGCAAGATTTGGACACAACTTTGCACCAACATATGAATGGTT 726
QY 284 LysArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
DB 727 AAAGACAGAGCATATTATACCGAAGATATAPACCAGTCTCAGCTGATGCGAGTATGAT 786
QY 303 GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn 322
DB 787 ATAAGTCTCAGCGATGATGTTCTCTCTATATTTGGGTACAAATCTGCTAATATAAT 846
QY 323 PheThrPheThrGlu-----SerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
DB 847 TTTACCTATGTGGATGAGAACTCTTGAGGTCACT-----TCAAAAGCAATCAGC 897
QY 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
DB 898 CAACGTGTGCTGATCTCATCCATTTTGGGAGAAGTTCGCAAAAGCTCCTGAGGCTTCT 957
QY 361 LeuGluSerLysGluAlaGlnLysLeuLysLeuAspGluLysAsnHisArgLysGlnIle 380
DB 958 ACCAGGAAGATGCTGCTCAGAACAAATTTCTGGAAGTAACTGCTCAGAAATGCATATA 1017
QY 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
DB 1018 GACACAGTGTGAAATATTATTTGGGAGTCTTTTATTTGGCATTTGAAAGGCTCCAGAGTA 1077
QY 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
DB 1078 CTCATATGCTGTAGACCGCTGGAATGGCACTTTGTCATGACTGGGACTGCCTGAAATA 1137
QY 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
DB 1138 ATGGTAAGGACTTTTGAGACATATTTGTGGATCTTGTCTCAGTATGGATGAATATATG 1197
QY 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
DB 1198 AGTCTCTTTGCCAAACATCTGCAATGCAAGATTAAGAATGACCAATGCTGATGCCTCA 1257
QY 461 GluGlnAlaCys 464
DB 1258 GCACAGCTTGT 1269

RESULT 10
AAQ50559
ID AAQ50559 standard; cDNA to mRNA; 1323 BP.
XX
AC AAQ50559;
XX
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase sequence (1).
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
JP05276960-A.
XX
PD 26-OCT-1993.
XX
PF 07-AUG-1992; 92JP-0231602.
XX
PR 07-FEB-1992; 92JP-0056023.
XX
SY (SYUZ/) SYUZO T.
XX
WPI; 1993-373587/47.
XX
New gene for encoding asparaginyl endo-peptidase - comprises 8

```

```

PT specified DNA sequences
XX Claim 1; Page 12; 35pp; Japanese.
XX
A gene encoding asparaginylendopeptidase is claimed.
8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
derived from a seed of Canavalia ensiformis which selectively
hydrolyses C-terminus amide bond of L-asparagine residue
(see AAR43033 and AAR43041).
The enzyme is useful for protein fragmentation and enzymatic
peptide synthesis.
The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
by PCR.
XX
SQ Sequence 1323 BP; 385 A; 246 C; 346 G; 346 T; 0 other;
XX
Alignment Scores:
Pred No.: 2,18e-96 Length: 1323
Score: 1080.00 Matches: 207
Percent Similarity: 65.278 Conservative: 73
Best Local Similarity: 48.258 Mismatches: 129
Query Match: 44.018 Indels: 20
DB: 14 Gaps: 4
XX
US-09-934-066-2 (1-466) x AAQ50559 (1-1323)
QY 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
DB 7 GGAACCCGGTGGCGGTTCTCGGGCTGTTCAACCGTTATCGAAATTTATAGCGATCAA 66
QY 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
DB 67 CGCGATGTGCCATGCTTACCAGTTGCTCATAAAAGTGGAGTCAAGAGCAAAACATT 126
QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
DB 127 GTGCTGTTTATGATGATATAGCTTATAGCCCATGAATCCCGAGCCGAGTCATC 186
QY 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
DB 187 ATCAACCATCTCAGGGCCAGACGCTGATGCTGTGTACCTAACGNTTACACCGCTAG 246
QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
DB 247 GACGTAAACCTGAGAACCTATATGCTGTCATCTTGGGACAAAGATAAAGTTAAGGT 306
QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
DB 307 GGAAGTGGCAAGGTGATCAACAGCAATCCCGAGGATAGGATATTTATATCTACTCCGAT 366
QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
DB 367 CATGGAGTCCCGAGTTCTTGGGATCCCAACGCCACCATTCGTTTATGCCATGATTTT 426
QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
DB 427 ATTGATGTTTGAAGAAGAACATGCAAGTGGAGGCTACAAGGAGATGCTTATACATA 486
QY 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
DB 487 GAAGCTTGTGAGAGTGGAGCATATTTGAGGTATTAATGCCCAAGGATCTGAATATTAT 546
QY 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
DB 547 GTGACAACTCGCTCAATGCACAGAACAGATTTTGGAACTTATTTGCTCGGGATGAAT 606
QY 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
DB 607 CCTCCTCCACAGAGAGTAGTACGTAACCTGCCCTGGGGGATTTATACAGCCTTTCTTTGGATG 666
QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
DB 667

```

```

Db 667 GAAGATAGTGAGACTACAATCTAAAAGGGAACGGTACACAGCAATACCACTCGGTA 726
Qy 284 LysArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
Db 727 AGGAACGGACTTCAAAATCTAACAGCTATAGTTGGTCTCATGTGATGCAATACGGT 786
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrLeuArgAsnProGluAsnAsp 321
Db 787 GACACTAACATTAAGTCTGCTGAAAAGCTTTACTTGTACCATGCTTTGTATCTCTGCCACCGTG 846
Qy 322 AsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeu----- 338
Db 847 AACCTT-----CCTCCACACACAGCGCAACCTAGAGCTAAA 882
Qy 339 -----ValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
Db 883 ATGGAAGCTTGTAAACAGAGAGATGCAGAGCTTTTGTTCATGTGCGCAATGTATCAGAGA 942
Qy 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
Db 943 TCAAAACCATCAACCGGAAAAGAGACTCACATCTCTGGAACAGATTACAGAGACAGTGAAG 1002
Qy 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
Db 1003 CATAGGAATCATTTGGATGGCAGTGTGGAATTGATTGGAGTTTGTGTATGGACACGGA 1062
Qy 396 AsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTyr 415
Db 1063 AAAAGTTCTTCGGTCTACATTCCGTGAGGCTCCTCGTCTGCCCTAGTTGATGATTGG 1122
Qy 416 AspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyr 435
Db 1123 ACATGCTTGAATCTATGTTAGAGTGTTCGAAACTCACTCTGGGTCACTGACTCAGTAT 1182
Qy 436 GlyLeuLysThrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln 455
Db 1183 GGCATGAACACATCGCGGATTCGCGAACGTTTTCGCAACGTCGCGGTT----- 1230
Qy 456 ThrValSerAlaIleGluGlnAlaCys 464
Db 1231 TCTAAGGCCCTCCATGGAGGAGGCTTGT 1257

RESULT 11
AAQ50563
ID AAQ50563 standard; cDNA to mRNA; 1323 BP.
XX
AC AAQ50563;
XX
DT 24-MAY-1994 (first entry)
XX
DE Asparaginylendopeptidase sequence (5).
XX
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX
OS Canavalia ensiformis.
XX
JN JP05276960-A.
XX
PD 26-OCT-1993.
XX
PF 07-AUG-1992; 92JP-0231602.
XX
PR 07-FEB-1992; 92JP-0056023.
XX
PA (STUZ/) SYUZO T.
XX
DR WPI; 1993-373587/47.
XX
PT New gene for encoding asparaginyl endo-peptidase - comprises 8
XX specified DNA sequences

```

```

PS Claim 1; Page 14; 35pp; Japanese.
XX
CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AAR43033 and AAR43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX
SQ Sequence 1323 BP; 383 A; 247 C; 346 G; 347 T; 0 other:

Alignment Scores:
Pred. No.: 2.18e-96 Length: 1323
Score: 1080.00 Matches: 207
Percent Similarity: 65.27% Conservative: 73
Best Local Similarity: 48.25% Mismatches: 129
Query Match: 44.01% Indels: 20
DB: 14 Gaps: 4

US-09-934-066-2 (1-466) x AAQ50563 (1-1323)
Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 7 GGTACCCTGGTGGCGGTTCTCGTGGCTGTCAACCGTTATGGAATTTATAGGCATCAA 66
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 67 GCGGATGTGTCCATGCTTACCAGTTGTATATAAAGGTGGAGTCAAGAGAAACATTT 126
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 127 GTGTGTATTATGATGATATAGCTTATAACCCCATGAATCCAGACCCGGAGTCACTC 186
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 187 ATCAACCATCTCAGGGCCAGACGCTGTATGCTGTACCTAAGGATTACACCGGTGAG 246
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 247 GACGTAACACCTGAGAACCTTATATGCTGTCTATCTTGGGACAGAGATAAAGTTAAAGT 306
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 307 GGAAGTGGCAAGGTGATCAACAGCAATCCGAGGATAGGATATTTATATTCTACTCCGAT 366
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 367 CATGGAGGTCCCGGAGTCTTGGGATGCCAACCCACCATTCGTTTACGCCCATGGATTTT 426
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 427 ATTGATGTTTTGAAGAAGAACATGCAAGTGGAGGCTACAAGGAGATGGTTTATATACATA 486
Qy 204 GluAlaCysGluSerGlySerMetPheGluClyLeuLysLysAsnLeuAsnIleTyr 223
Db 487 GAAGCTTGTGAGTGGGAGCATATTTGAGGATATAATGCCCAAGGATCTGAATATTTAT 546
Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 547 GTGACAACTGCGTCAATGCACAAGACAGACAGTTTGTGAACTTTATTCCTCGGTGAAT 606
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 607 CCTCTCCACAGAAAGAGTAGTACGTAACCTTGGCTGGGGATTTATACAGCGTTCTTGGATG 666
Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
Db 667 GAAGATAGTGAGACTACAATCTAAAAGGGAACGGTACACAGCAATACCACTCGGTA 726

```



```

Db 895 GAAGATAGTGAGACTCAATCTAAAAAGGGAACGGTACACAGCAATACCAAGTCGGTA 954
Qy 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
Db 955 AGAAGAGGACTCAAAATCTAACAGCTATAGGTTGGTTCTCTATGTGATGCAATACGGT 1014
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321
Db 1015 GACACTAACATTACTGCTGAAAAAGCTTTACTGTGACCATTGTTGTATCTCTGCCACCGTG 1074
Qy 322 AsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeu----- 338
Db 1075 AACTTT-----CCTCCACACACAGCGCAACCTAGAGAGCTAAA 1110
Qy 339 -----ValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
Db 1111 ATGGAGTTGTTAACCCAGAGAGTGCAGAGCTTTTGTTCATGTGGCAATGTATCAGAGA 1170
Qy 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
Db 1171 TCAACCATCAACCGGAAAAAGACTCACATCTCTGGAACAGATTACAGAGACAGTGAAG 1230
Qy 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
Db 1231 CATAGGAATCATTTGGATGCGACGTGTGGAATGATTGGAGCTTTTGTGTATGGACACGGA 1290
Qy 396 AsnValLysAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyr 415
Db 1291 AAAAGTCTTCGGTTCTACATCTCGTGAGGCGCTCGTCTGCCCTAGTTGATGTTGG 1350
Qy 416 AspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyr 435
Db 1351 ACATGCTTGAATCTATGTTAGAGTGTTCGAAACTCACTGTGGTCACTGACTCACTAT 1410
Qy 436 GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln 455
Db 1411 GGATGAACACATCGCGGCTATTCGGCAACGTTTGGCAACAGCGCGGTT----- 1458
Qy 456 ThrValSerAlaIleGluGlnAlaCys 464
Db 1459 TCTAAGGCCCTCCATGGAGAGGCTGT 1485

```

RESULT 13

AAQ50575

ID AAQ50575 standard; cDNA to mRNA; 1640 BP.

AC AAQ50575;

XX AAQ50575;

XX 24-MAY-1994 (first entry)

XX Asparaginylendopeptidase clone ASN-1.

XX Asparaginylendopeptidase; Canavalia ensiformis; seed;

XX L-asparagine; primer; PCR; protein fragmentation;

XX peptide synthesis; ss.

XX Canavalia ensiformis.

XX JP05276960-A.

XX 26-OCT-1993.

XX 07-AUG-1992; 92JP-0231602.

XX 07-FEB-1992; 92JP-0056023.

XX (SYUZU) SYUZU T.

XX WPI; 1993-373587/47.

XX New gene for encoding asparaginyl endo-peptidase - comprises 8

XX specified DNA sequences

xx Disclosure; Page 26; 35pp; Japanese.

xx A gene encoding asparaginylendopeptidase is claimed.

cc 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease

cc derived from a seed of Canavalia ensiformis which selectively

cc hydrolyses C-terminus amide bond of L-asparagine residue

cc (see AAR43033 and AAR43041).

cc The enzyme is useful for protein fragmentation and enzymatic

cc peptide synthesis.

cc The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used

cc in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,

cc by PCR.

xx Sequence 1640 BP; 533 A; 288 C; 383 G; 436 T; 0 other;

Alignment Scores:

Pred. No.: 2.88e-90 Length: 1640

Score: 1019.00 Matches: 195

Percent Similarity: 64.42% Conservative: 73

Best Local Similarity: 46.88% Mismatches: 128

Query Match: 41.52% Indels: 20

DB: 14 Gaps: 4

US-09-934-066-2 (1-466) x AAQ50575 (1-1640)

Qy 57 TyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGly 76

Db 4 TATGGTAATTATCGGCATCAAGCGGATGTGCCATGCTTACCAGTTGCTGATAAAGGT 63

Qy 77 GlyLeuLysAspGluAsnIleIleValPheMetTyrAspIleAlaPheSerSerGlu 96

Db 64 GGAGTCAAGAAGAAACATTTGGTGTATGTATGATGATATAGCTTATAACGCCATG 123

Qy 97 AsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysLysVal 116

Db 124 ATCCAGACCGGAGTCATCATCAACCATCTCCAGGGCCGACGCTGATGCTGGTGA 183

Qy 117 ProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLysGly 136

Db 184 CCTAAGGATTACACCGGTGAGGAGGTAAACACCTCAGAACCTATATGCTCATCTTGGG 243

Qy 137 AsnGluSerGlyValThrGlyLysValLysValLysValLysValLysValLysVal 156

Db 244 GACAAGAGTAAAGTTAAAGGTGGAAGTGGCAAGGTGATCAACAGCAATCCGAGGATAGG 303

Qy 157 IlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAsp 176

Db 304 ATATTATTTCTACTCCGATCAGGAGGTCCCGAGTTCTTGGGATGCCAAGCGCACCA 363

Qy 177 GluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyr 196

Db 364 TTCGTTTATGCCATGGATTTTATTGATGTTTGAAGAAGAAACATGCAAGTGGAGGTAC 423

Qy 197 AsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyLeuLeu 216

Db 424 AAGGAGATGTTATATACATAGAACGTTGTGAGAGTGGGAGCATATTTGAGGGTATATG 483

Qy 217 LysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGly 236

Db 484 CCCNAGGATCTGAATATTTATGTGACACCTCCGTCAAATGCCAAGAGACAGTTTTTGA 543

Qy 237 ValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLysLysAsp 256

Db 544 ACTTATTGTCCTGGGATGAATCTCTCCCTCCACAGAGATACGTAACTTGCCTGGGGAT 603

Qy 257 ThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeu 276

Db 604 TTATACAGCGTTTCTTGATGGAAGATAGTGAGACTCACAACTCAAAAAGGAAACGGTA 663

Qy 277 GluGlnGlnTyrHisValValLysArgValGlySerAspValProGluThr----- 294

Db 664 CAACAGCAATACCAGTCGGTAAGGAACGGACTTCAAAATCTTAACAGCTATAGGTTGGT 723

295 SerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIle 314
 724 TCTCATGTGATGCAATACGGTCACACTACATCTACTGCTGAAAGCTTACTTGTACCAT 783
 315 GlyArgAsnProGluAsnAspAsnPheThrGluSerPheSerSerProIleSer 334
 784 GGTTCATGCTCCACCGCTGAACCTT-----CCCTCCACAC 819
 335 AsnSerGlyLeu-----ValAsnProArgAspIleProLeuLeuTyr 348
 820 AACGGCAACCTAGAACCTAAATGGAAGTGTGTAACACGAGAGATGCGAGCTTTTGTTC 879
 349 LeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLys 368
 880 ATGTGGCAAAATGATCAGAGATCAACACCATCAACCGGAAAGAAAGACACTCACATCTGGAA 939
 369 LysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeu 388
 940 CAGATTACAGACAGACTGAAGCATAGGAATCATTTGGATGGCAGTGTGGAATTGATTGGA 999
 389 ArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrGly 408
 1000 GTTTTGTGTGATGGACCGAGAAAGTCTCTCGGTCTACATTCGCTGAGGCTCCTGTT 1059
 409 GlnProLeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHis 428
 1060 CTGCCCTAGTGTGATGGACATGCTTGAATCTAGTGTAGAGTGTTCGAACACTCAC 1119
 429 CysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsn 448
 1120 TGTGGGTCACTGACTCAGTATGGCATGAAACACATCGCGGGCATTCGCGAACGTTTGGCAAC 1179
 449 MetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
 1180 AGCGGGCTT-----TCTAAGCGCTTCCATGGAGGAGGCTTGT 1215
 RESULT 14
 ID AAQ50564 standard; cDNA to mRNA; 1152 BP.
 AC AAQ50564;
 XX 24-MAY-1994 (first entry)
 DT
 DE Asparaginyl endopeptidase sequence (6).
 KW Asparaginyl endopeptidase; Canavalia ensiformis; seed;
 KW L-asparagine; primer; PCR; protein fragmentation;
 KW peptide synthesis; ss.
 XX Canavalia ensiformis.
 XX JP05276960-A.
 XX 26-OCT-1993.
 XX 07-AUG-1992; 92JP-0231602.
 XX 07-FEB-1992; 92JP-0056023.
 XX (SYUZU) SYUZU T.
 XX WPI; 1993-373587/47.
 XX New gene for encoding asparaginyl endo-peptidase - comprises 8
 XX specified DNA sequences
 PS Claim 1; Page 14-15; 35pp; Japanese.
 CC A gene encoding asparaginyl endopeptidase is claimed.
 CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
 CC derived from a seed of Canavalia ensiformis which selectively

hydrolyses C-terminus amide bond of L-asparagine residue
 (see AAQ3033 and AAQ3041).
 The enzyme is useful for protein fragmentation and enzymatic
 peptide synthesis.
 The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
 in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
 by PCR.
 XX Sequence 1152 BP; 340 A; 210 C; 297 G; 305 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,266-87 Length: 1152
 Score: 990.00 Matches: 192
 Percent Similarity: 65.38% Conservative: 63
 Best Local Similarity: 49.23% Mismatches: 119
 Query Match: 40.34% Indels: 16
 DB: 14 Gaps: 3
 US-09-934-066-2 (1-466) x AAQ50564 (1-1152)
 QY 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
 DB 7 GGTACCCGGTGGCGGTTCTCGTGGCTGTTCAACAGGTTATGCAAAATTTATAGCGATCAA 66
 QY 64 AlaAspIleCysHisAlaTyrClnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
 DB 67 GCGATGTGTGGCATGCTTACCGATTGCTGATAAAGGTGGAGTCAAGAAGAACAACAT 126
 QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
 DB 127 GTGGTGTATATGATGATGATATAGCTTATAACGCCATGAATCCAGACCCGAGTGCATC 186
 QY 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
 DB 187 ATCAACCATCTCTCAGGGCCAGACGCTGATGCTGGTACCTAAGGATTACACCGGTGAG 246
 QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
 DB 247 GACGTAACACCTGAGAACCTATATGCTCTCTCTTGGGACAGAGTAAAGTTAAAGGT 306
 QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
 DB 307 GGAAGTGCGAAGTGTATCAACAGCAATCCGGAGGATAGGATATTTATATTCTACTCCGAT 366
 QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
 DB 367 CATGGAGTCCCGGAGTCTTGGGATGCCAACGCCACCATTCGTTTACGCCATGGATTTT 426
 QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
 DB 427 ATTGATGTTTGAAGAAGAACAATCCAACTGCGAGCTCAAGAGAGTGGTTATATACATA 486
 QY 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
 DB 487 GAAGCTTGTGAGAGTGGGAGCATATTTGAGGGTATAATGCCCAAGGATCTGAATATTTAT 546
 QY 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
 DB 547 GTCACAACTCGCTCAAAATGCACAGAGAACAGTCTTGGAACTTATTCCTCTGGGATGAAT 606
 QY 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
 DB 607 CCTCTCCACCAAGAGAGTACCTAACTTGCCTGGGGATTTTATACAGCGTTCTTTCGATG 666
 QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
 DB 667 GAAGATAGTACACTCACAATCTAAAGGCAACGCTACACAGCAATACCAGTCCGTA 726
 QY 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
 DB 727 AGCAACGCACTTCAAAATCTTAACAGCTATAGTTTGTCTCATGTGATGCAATACCGT 786
 QY 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321


```
Db 787 GACACTAATCTACTGCTGAAAGGCTTACTTGTACCATGGTTTGTGATCTCTCCACCGTG 846
Qy 322 AsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeu----- 338
Db 847 AACTTT-----CCTCCACACACGGCAACCTAGAGCTAAA 882
Qy 339 -----ValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
Db 883 ATGGAAGTTGTTAACACAGAGATGCCAGAGCTTTTGTTCATGTGCGCAATGTATCAGAGA 942
Qy 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
Db 943 TCACCAACCATCAACCGGAAAGAGACTCACATCTCTGGAACAGATTACAGAGACAGTGAAG 1002
Qy 376 HisArgLysGlnIleAspGlnSerIlePheThrAspIleLeuArgLeuSerValLysGlnThr 395
Db 1003 CATAGGAATCATTTGGATGGCAGGTGTGGAATGTATGGAGTTTGTTCATGACAGGAC 1062
Qy 396 AsnValLeuAsnLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrp 415
Db 1063 AAAAGTTCTCGGTTCTACATTCCTGAGGGCTCTGCTGCTGCCCTAGTTGATGATTGG 1122
Qy 416 AspCysPheLysThrLeuValAsnSerPhe 425
Db 1123 ACATGCTTGAATCTATGTTAGAGTGTTTC 1152

RESULT 15
ID AAQ50565 standard; cDNA to mRNA; 894 BP.
XX AC AAQ50565;
XX DT
XX DE
XX DE Asparaginylendopeptidase sequence (7).
KW Asparaginylendopeptidase; Canavalia ensiformis; seed;
KW L-asparagine; primer; PCR; protein fragmentation;
KW peptide synthesis; ss.
XX OS Canavalia ensiformis.
XX PN JP05276960-A.
XX PD
XX PR 26-OCT-1993.
XX PF 07-AUG-1992; 92JP-0231602.
XX PR 07-FEB-1992; 92JP-0056023.
XX PA (SYUZ/) SYUZU T.
XX PA WPI; 1993-373587/47.
DR
XX New gene for encoding asparaginyl endo-peptidase - comprises 8
XX specified DNA sequences
PS Claim 1; Page 15; 35pp; Japanese.
XX
CC A gene encoding asparaginylendopeptidase is claimed.
CC 8 DNA sequences are given (AAQ50559-66). The enzyme is a protease
CC derived from a seed of Canavalia ensiformis which selectively
CC hydrolyses C-terminus amide bond of L-asparagine residue
CC (see AA43033 and AA43041).
CC The enzyme is useful for protein fragmentation and enzymatic
CC peptide synthesis.
CC The primers given in AAQ50567-68, AAQ50576-77 and AAQ50583-90 were used
CC in the isolation of the fragments given in AAQ50569-75 and AAQ50578-79,
CC by PCR.
XX
SQ Sequence 894 BP; 265 A; 166 C; 229 G; 234 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 4.77e-78 Length: 894
Score: 891.00 Matches: 162
Percent Similarity: 74.11% Conservative: 47
Best local Similarity: 57.45% Mismatches: 71
Query Match: 36.31% Indels: 2
DB: 14 Gaps: 1
US-09-934-066-2 (1-466) x AAQ50565 (1-894)
Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 7 GGTACCGGTGGGGGTTCTCGTCTGTTCAACCGTTATGGAATATTAGGATCAA 66
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 67 CGGATGTGTGCCATGCTTACCAGTGTCTGATAAAGGTGGAGTCAAGAGAGAAACAT 126
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 127 GTGTGTGTTATGTATGATGATATAGCTTATAACGCCATGAATCCAGACCCGGAGTCATC 186
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 187 ATCAACCATCTCTAGGGGCCAGACGTGTATGCTGTGTACCTAAGGATTACACCGGTGAG 246
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 247 GACGTACACCTGAGAACCTATATGCTGTCTATCTTGGGGACAGAGTAAGTTAAAGGT 306
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 307 GGAAGTGGCAAGTGCATCAACAGCAATCCGGAGGATAGGATATTTATATTCTACTCCGAT 366
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 367 CATGGAGTCCCGGAGTTCTTGGGATGCCAACCCACCATTCCTGTACGCCATGGATTTT 426
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 427 ATTGATGTTTGAAGAGAAACATGCAAGTGGAGGCTACAAGGAGATGGTTATATACATA 486
Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 487 GAAGCTTGAGAGTGGGAGCATATTTGAGGGTATATATGCCCAAGGATCTGAATATTTAT 546
Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 547 GTGACAACTGCTCAAAATGCACAAGAGAACAGTTTGTGGAACCTATTGTCTCGGATGAAT 606
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 607 CCTCTCCACAGAGAGTACGTACTTGCCTGGGGATTTATACAGCGTTCTTGGGATG 666
Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValVal 283
Db 667 GAAGATACTGAGACTCACAAATCTAAAAGGGAACGGTACACAGCANTACACGTCGCTA 726
Qy 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArgPheGly 301
Db 727 AGGAACCGGACTTCAAAATCTAACAGCTATATAGTGTGTGTTGTTCTCATGTGCAATACG 786
Qy 302 ThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAsp 321
Db 787 GACACTAACATTACTGCTGAAAGCTTTTACTTGTACCATGGTTTGTGATCTCTCCACCG 846
Qy 322 AsnPhe 323
Db 847 AACTTT 852
```

Search completed: May 29, 2003, 18:27:15
Job time : 329 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:17:59 ; Search time 82 Seconds
(without alignments)
1742.822 Million cell updates/sec

Title: us-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHFOLVFLHALLIF.....CNGMVQDVQKTVSAIEQACSM 466

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09934066/runat_19052003_163700_24409/app_query.fasta_1.647
-DB-Issued_Patents_NA -QPMF=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosu62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09934066.cgn_1_1_61_@runat_19052003_163700_24409 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	828	33.7	1936	2	US-09-139-424-3
2	828	33.7	2030	3	US-08-706-216-3
3	744	30.3	1855	2	US-08-928-613-1
4	744	30.3	1855	3	US-09-193-524-1
5	178	7.3	220	2	US-08-928-613-15
6	178	7.3	220	3	US-09-193-524-15
7	168	6.8	219	2	US-08-928-613-11
8	168	6.8	219	3	US-09-193-524-11
9	137	5.6	1860	4	US-09-134-001C-1411
10	136	5.5	544	4	US-09-280-116-247
11	135.5	5.5	155	2	US-08-928-613-5
12	135.5	5.5	155	3	US-09-193-524-5

13	116	4.7	229	2	US-08-928-613-13	Sequence 13, Appl
14	116	4.7	229	3	US-09-193-524-13	Sequence 13, Appl
15	115.5	4.7	9510	4	US-09-453-702B-256	Sequence 256, App
16	114.5	4.7	331	2	US-08-928-613-10	Sequence 10, Appl
17	114.5	4.7	331	3	US-09-193-524-10	Sequence 10, Appl
18	114.5	4.7	1743	4	US-09-367-583-1	Sequence 1, Appl
19	112.5	4.6	4850	4	US-09-443-184-34	Sequence 34, Appl
20	104.5	4.3	206	2	US-08-928-613-22	Sequence 22, Appl
21	104.5	4.3	206	3	US-09-193-524-22	Sequence 22, Appl
22	103.5	4.2	199	2	US-08-928-613-18	Sequence 18, Appl
23	103.5	4.2	199	3	US-09-193-524-18	Sequence 18, Appl
24	103.5	4.2	13121	4	US-08-961-527-126	Sequence 126, App
25	103	4.2	4031	1	US-08-471-033-49	Sequence 49, Appl
26	103	4.2	4031	2	US-08-471-044-49	Sequence 49, Appl
27	103	4.2	4031	2	US-08-463-483A-49	Sequence 49, Appl
28	103	4.2	4031	2	US-08-471-046A-49	Sequence 49, Appl
29	103	4.2	4031	2	US-08-470-566B-49	Sequence 49, Appl
30	103	4.2	4031	2	US-08-469-334-49	Sequence 49, Appl
31	103	4.2	4031	3	US-09-300-529-49	Sequence 49, Appl
32	103	4.2	12687	1	US-08-676-169-1	Sequence 1, Appl
33	103	4.2	12687	3	US-08-981-459-1	Sequence 1, Appl
34	103	4.2	12687	4	US-09-063-431A-1	Sequence 1, Appl
35	102.5	4.2	11464	4	US-08-991-840A-2	Sequence 2, Appl
36	100.5	4.1	11459	4	US-09-454-721A-3	Sequence 3, Appl
37	100	4.1	1791	1	US-08-245-294-7	Sequence 7, Appl
38	100	4.1	1791	1	US-08-474-499-7	Sequence 7, Appl
39	100	4.1	1791	1	US-08-307-279A-7	Sequence 7, Appl
40	100	4.1	1791	4	US-09-525-310-7	Sequence 7, Appl
41	100	4.1	1791	5	PCT-US95-06211-7	Sequence 7, Appl
42	100	4.1	2877	4	US-09-619-353-1	Sequence 1, Appl
43	99.5	4.1	3792	2	US-08-992-334-1	Sequence 1, Appl
44	99.5	4.1	3792	3	US-08-302-752-1	Sequence 1, Appl
45	99.5	4.1	5234	2	US-08-992-334-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-139-424-3
; Sequence 3, Application US/09139424
; Patent No. 5985832
; GENERAL INFORMATION:
; APPLICANT: Roodman, G. David
; APPLICANT: Reddy, Sakamuri V.
; APPLICANT: Choi, Sun-Jin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE FOR
; TITLE OF INVENTION: OSTEOCLAST INHIBITORY FACTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,424
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/772,441
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSK:295
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1936 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 136..1434

US-09-139-424-3

Alignment Scores:

Pred. No.: 9,15e-91 Length: 1936
Score: 828.00 Matches: 179
Percent Similarity: 55.34% Conservative: 80
Best Local Similarity: 38.25% Mismatches: 157
Query Match: 33.74% Indels: 52
DB: 2 Gaps: 13

US-09-934-066-2 (1-466) x US-09-139-424-3 (1-1936)

Qy 8 PheGlnLeuValPheLeuHisAlaLeuLeuLeuPheSerAlaGluSerArgLysThr 27
Db 142 TGAAGAGTACGTATCTCTCAGTGTGGCCCTGGGCATTTGGTGC - - - - - 186
Qy 28 GlnLeuLeuAsnAspAsnAspValGluSerSerLysSerAlaLysGlyThrArgTrp 47
Db 187 - - - - - GTTCTATAGATGATCTCTGAGATGAGATGAGGCAAGCACTGG 225
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 226 GTGGTATGTCGTCAGTTCATCAATGCTGATATATATAGCCACGAGCAGCGCTGC 285
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleValPheMet 87
Db 286 CATGCCCTACCATCATTCACCGCAATGGATTCCTGACGAAACAGATGCTGTGATGATG 345
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 346 TACGATGACATGCTTACTCTCAAGACAATCCCATCCAGCAATGATGATCAACAGGCC 405
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db 406 AATGGCACAGATGCTATCAGGAGTCCCGAAGACTACATCGAGGAGGATTTACCCCA 465
Qy 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly - - - GlyAsnGly 146
Db 466 CAATATTCCTGCTGTGTGAGCGGATCGAGAGCAGTGAAGGCGATAGGATCCGCG 525
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheLeuTyrTyrAlaAspHisGlyAla 166
Db 526 AAGTCTGAAGAGTGGCCCGCAGGATCAGCGTGTTCATTTACTTCACTGACCATGGATCT 585
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 586 ACTGGAATCTGTTTTTCC - - - AATGAAGATCTTCATGTAAAGGACCTGAATGAGACC 642
Qy 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 643 ATCCATTACATACAAACACAAATGTACCGAAAGATGGTGTCTACATTGAAGCCTGT 702
Qy 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 703 GAGTCTGGTCCATGATGAACCAC - - - CTGCGGATACATCAATGTTATGCAACTACT 759
Qy 227 AlaAlaAsnSerLysLysSerTrpGlyValTyrCysProGluSerTyrProProPro 246
Db 760 GCTGCCAACCCAGAGATGCTCTACCGCTGTACTATGATGAG - - - - - 804
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 805 - - - - - AAGAGTCCACGATCTGCGGAGCTGGTACACGCTCAACTGGATGGAAGACTCG 858

Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnIleTyrHisValValLysArgArg 286
Db 859 GACGTGAAGATCTGACTAAAGAGACCTCGACAGCAGTACCACTGGTAAATCGCAC 918
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 919 ACC - - - - - AACACCAAGCCACCTCATGAGTATGCAACAAACAAATCTCC 963
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 964 ACCATGAAGTGCATCAGTTTCAGGT - - - - - ATGAAA 996
Qy 327 GluSerPheSerSer - - - - - ProIleSerAsnSerGlyLeuValAsnPro 341
Db 997 CGCAAAAGCCAGTCTCCGCTCCCTACCTCCAGTCACACACCTTGACCTACCCCGCAGC 1056
Qy 342 ArgAspIleProIleLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 1057 CCTGATGCTCTCTCCATCATGAAAGAACTGATGAACACC - - - - - AATGATCTG 1110
Qy 362 - - - - - GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1111 GAGGAGTCCAGCAGCTCCAGGAGAGATCCAGCGGATCTGGATGCCAGGACCTCATTT 1170
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1171 GAGAAGTCAGTCGTAAGATCTCTCTTGTGTCGAGGCTCCGAGGCTGAGGAGCAG 1230
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCysPheLysThr 420
Db 1231 CTCTGTCGACAGAGCC - - - - - CCGCTCACGGGGCAC - - - - - AGCTCTACCCAGAG 1278
Qy 421 LeuValAsnSerPheLysAsnHisCys - - - - - GlyAlaThrValHisTyrGly 436
Db 1279 GCCTGCTGCACTTCGGGACCCACTGCTTCACTGCGACCTCCCGCAGTACGATGCG 1338
Qy 437 LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr 456
Db 1339 TTGAGACATTTGACGCTGCTGTCACCTTTGTGAGAAGCCGTATCCGCTTCACAGGATA 1398
Qy 457 ValSerAlaIleGluGlnAlaCys 464
Db 1399 AATTTGTCATGGACCCAGCTGTGC 1422

RESULT 2

US-08-706-216-3

Sequence 3, Application US/08706216

Patent No. 6140098

GENERAL INFORMATION:

APPLICANT: Balasubramanian, Sri Ram

APPLICANT: Ford, John

APPLICANT: Gorman, Daniel M.

APPLICANT: Zurawski, Gerard

TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: DNAX Research Institute

STREET: 901 California Avenue

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/706.216

FILING DATE: 30-AUG-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ching, Edwin P.					
REGISTRATION NUMBER: 34,090					
REFERENCE/DOCKET NUMBER: DX0613					
TELECOMMUNICATION INFORMATION:					
TELEPHONE: 415-852-9196					
TELEFAX: 415-496-1200					
INFORMATION FOR SEQ ID NO: 3:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 2030 base pairs					
TYPE: nucleic acid					
STRANDEDNESS: single					
TOPOLOGY: linear					
MOLECULE TYPE: cDNA					
FEATURE:					
NAME/KEY: CDS					
LOCATION: 183..1484.					
US-08-706-216-3					
Alignment Scores:					
Pred. No.: 9,9e-91 Length: 2030					
Score: 828.00 Matches: 179					
Percent Similarity: 55.34% Conservative: 80					
Best Local Similarity: 38.25% Mismatches: 157					
Query Match: 33.74% Indels: 52					
DB: Gaps: 13					
US-09-934-066-2 (1-466) x US-08-706-216-3 (1-2030)					
QY	8	PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr	27		
DB	189	TGAAGTAGCTGTATTCTCAGTGGCCCTGGGCATTTGGTGCC-----	233		
QY	28	GlnLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrp	47		
DB	234	-----GTTCTATAGATGTCCTGAGATGGAGCGAACACTGG	272		
QY	48	AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys	67		
DB	273	GTGGTGATCGTGGCAGGTCAAAATGGCTGGTATAATATTAGGCCACCGAGCGCGTGC	332		
QY	68	HisAlaTyrglileLeuargLysglyLeuLysAspGluAsnIleIleValPheMet	87		
DB	333	CATGCCCTACCAGATCATTCACCGCAATGGGATCTCTGACGAACAGATCGTTGTGATGATG	392		
QY	88	TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro	107		
DB	393	TAGCATGACATTGCTTACTCTGAACACATCCCACCTCCAGGAATGTGATCAACAGGCC	452		
QY	108	AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal	127		
DB	453	AATGGCACAGATGTCTATCAGGGAGTCCGGAAGACTACACTGGAGAGGATGTTACCCCA	512		
QY	128	GlnAsnPheTyrAsnValLeuLeuGluAsnGluSerGlyValThrGly--GlyAsnGly	146		
DB	513	CAAAATTTCTTGTGTGTGTGAGAGCGCATGACAGAGCAGTGAAGGGCATAGGATCCGGC	572		
QY	147	LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla	166		
DB	573	AAAGCTCTGAAGAGTGGCCCCAGCATCAGTGTTCATTACTTACTCAGTACCATGGATCT	632		
QY	167	ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal	186		
DB	633	ACTGGAATACTGGTTTTTCCC---AATGAAGATCTTCATGTAAAAGGACCTGAATGAGACC	689		
QY	187	LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys	206		
DB	690	ATCCCATATCATACAAACACAAAAATGACCGAAGATGGTGTCTACATTGAAGCGCTGT	749		
QY	207	GluSerGlySerMetPhecluglyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr	226		
DB	750	GAGTCTGGTCCATCATGATCAACAC---CTGCCGGATAACATCAATGTTTATGCAACTACT	806		
QY	227	AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro	246		

Db	807	GCTGCCAACCCAGAGAGTGCCTACGCCCTGTTACTATGATGAG-----	853
QY	247	ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer	266
DB	852	-----AAGAGGTCCACGTACCTGGGACTGTGTACAGCTCACTGGATGGAGACTCG	905
QY	267	AspLeuHisAspMetSerLysGluThrLeuGluGlnInTyrHisValValLysArgArg	286
DB	906	GACGTGGAAGATCTGACTAAAGAGAGCCCTGCACAAGCAGTACCACCTGGTAATAATCGCAC	965
QY	287	ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu	306
DB	966	ACC-----AACACCAGCCACGCTCATCGTAGTATGGAACAAACAATCTCC	1010
QY	307	LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr	326
DB	1011	ACCATGAAGTATGATGAGTTTCAGGT-----ATGAAA	1043
QY	327	GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro	341
DB	1044	CGCAAAGCCAGTTCTCCCGTCCCTACTCTCCAGTCACACACCTTGACCTCACCCCCAGC	1103
QY	342	ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu	361
DB	1104	CCTGATGTGCTCTCACCATCATGAAAGAACTGATGAACACC-----AATGATCTG	1157
QY	362	--GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle	380
DB	1158	GAGGAGTCCAGGAGCTCAGCGGAGAGATCCACGGCGCATCTGGATGCCAGGACGAGCAG	1217
QY	381	AspGlnSerIleThrAspIleLeuLeuLeuSerValLysGlnInThrAsnValLeuAsnLeu	400
DB	1218	GAGAAGTCAAGTCGGTAAGATCGTCTCTTGTGGCAGCGTCCGAGGCTGAGTGGAGCAG	1277
QY	401	LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr	420
DB	1278	CTCCTCTCCGAGAGAGCC-----CCGCTCACGGGGCAC---AGCTGTATCCACGAG	1325
QY	421	LeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrCly	436
DB	1326	GCCTGCTGCATCTCCGGACCCACTGCTTCACTCCCGCCAGCTACAGGATATCGC	1385
QY	437	LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr	456
DB	1386	TTGAGACATTTGTACGTGCTGTGCTCAACCTTTGTGAGAGAGCCGTATCCACTTCACAGGATA	1445
QY	457	ValSerAlaIleGluGlnAlaCys	464
DB	1446	AAATTGTCCATGGACCGTGTGC	1469
RESULT 3			
US-08-928-613-1			
Sequence 1, Application US/08928613			
Patent No. 5840562			
GENERAL INFORMATION:			
APPLICANT: Dieph, Dinh			
APPLICANT: Braxton, Scott M.			
APPLICANT: Delegeane, Angelo M.			
TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE			
NUMBER OF SEQUENCES: 24			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.			
STREET: 3174 Porter Drive			
CITY: Palo Alto			
STATE: CA			
COUNTRY: USA			
ZIP: 94304			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			

: CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/928,613
 : FILING DATE:
 : CLASSIFICATION: 524
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/567,506
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Luther, Barbara J.
 : REGISTRATION NUMBER: 33954
 : REFERENCE/DOCKET NUMBER: PF-0048 US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 415-855-0555
 : TELEFAX: 415-852-0195
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1855 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : IMMEDIATE SOURCE:
 : LIBRARY: Adrenal
 : CLONE: 100877
 : US-08-928-613-1

Alignment Scores:
 Pred. No.: 1,51e-80 Length: 1855
 Score: 744.00 Matches: 166
 Percent Similarity: 54.63% Conservative: 70
 Best Local Similarity: 38.43% Mismatches: 128
 Query Match: 30.32% Indels: 68
 DB: 2 Gaps: 14

US-09-934-066-2 (1-466) x US-08-928-613-1 (1-1855)

QY 8 PheGlnLeuValPheLeuHisAlaLeuLeuPheSerAlaGluSerArgLysThr 27
 DB 60 TGAAGAGTAGTGTATCTCTAGTGGCCCTGGGAATTGGTGCC----- 104
 QY 28 GlnLeuLeuAsnAspAsnValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
 DB 105 -----GTTCTATAGATGATCTGGAAGATGGAGGCAAGCACTGG 143
 QY 48 AlaValLeuValAlaGlySerAsnGlnTyrAsnTyrArgHisGlnAlaAspLeuCys 67
 DB 144 GTGGTGATCGTGGCAGGTTCAGATGGCTGGTATAATTATAGCAGCAGGCGCGTCG 203
 QY 68 HisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMet 87
 DB 204 CATGCTACCACTGATTCATCCAGCAATGGGATCTCTGCCAGACAGATCGTTGTGATTG 263
 QY 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
 DB 264 TAGATGATAGCTACTCTGAAGACATCCACTCCAGGAATGTGATCAACAGGCC 323
 QY 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
 DB 324 AATGGCAGAGTGTCTATCAGGAGTCCCGAAGGACTACACTGGAGGATGTATACCCCA 383
 QY 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
 DB 384 CAAATTTCTTGTGTGTGAGAGCGGATCCAGAACAGTGAAGGATAGGATCCCGC 443
 QY 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
 DB 444 AAGTCTGAAGAGTGGTCCCGCAGGATCACGTGTTCATTTATTTCTACCTGACCATGATCT 503
 QY 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
 DB 504 TCTGGAATACTGTTTCCCC---AATGAAGATCTTCATGTAAGGACCTGATTAAGACC 560
 QY 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206

DB 561 ACCATTATATTTTCAAAAACAAATGTACCGAAGATGGTGTCTACATTTGAGGCCGTGT 620
 QY 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
 DB 621 GAGTCTGGTCCATGATGAACAC---CTGCCGGAATAACATCAATGTTTATGCAACTACT 677
 QY 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro 246
 DB 678 GCTGCCAACCCAGAGAGCTGCTCTACCCCTGTTACTATGATGAG----- 722
 QY 247 ProSerGlyIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
 DB 723 -----AAGAGGTCCACCTACTCTGGGACCTGGTACAGCGTCAACTGGATGGAAGACTCG 776
 QY 267 AspLeuHisAspMetSerLysGluThrLeuGlnGlnTyrHisValValLysArgArg 286
 DB 777 GACGTGGAAGATCTGACTAAGAGACCTCTGCACAGCAGTACCACTGGTAAATCCAC 836
 QY 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
 DB 837 ACC-----AACACCACCCACCTGATGAGTATGGAACAAACAAATCTCC 881
 QY 307 LysAspTyrLeuSerSerTyrTleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
 DB 882 ACCATGAAGATGATGACAGTTTTCAGGGT-----ATGAAA 914
 QY 327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
 DB 915 CGCAAGCCAGTTCCTCCGCTCCCTACCTCCAGTCCACACCTTGACCTCAGCCCGCAGC 974
 QY 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
 DB 975 CCTGATGTGCTCTCACCATCATGAAAAGAACTGATGAACACC-----AATGATCTG 1028
 QY 362 GluSerLys-----GluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
 DB 1029 GAGGAGTCCAGCAGCTCACGAGGAGATCCAGCGGTATCTGGATGCCAGGCAC----- 1082
 QY 378 LysGlnIleAspGlnSerIleThrAspTleLeuArgLeuSerValLysGlnThrAsnVal 397
 DB 1083 -----CTCATCCGAGGTGAGGTGGAGCAG----- 1106
 QY 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCys 417
 DB 1107 -----CTCCTCTCCGAGAGACC-----CCGCTCAGCGGGCAC---AGCTGC 1145
 QY 418 PheLysThrLeuValAsnSerPheLysAsnHisCys 429
 DB 1146 TACCAGAGGTCTGTGTTACTTCCGACCCACTGC 1181

RESULT 4

: US-09-193-524-1
 : Sequence 1, Application US/09193524
 : Patent No. 6007997
 : GENERAL INFORMATION:
 : APPLICANT: Diep, Dinh
 : APPLICANT: Braxton, Scott M.
 : APPLICANT: Deleane, Angelo M.
 : TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 : STREET: 3174 Porter Drive
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/193,524
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/928,613
;; FILING DATE:
;; APPLICATION NUMBER: 08/567,506
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Luther, Barbara J.
;; REGISTRATION NUMBER: 33954
;; REFERENCE/DOCKET NUMBER: PF-0048 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-853-0555
;; TELEFAX: 415-852-0195
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1855 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; IMMEDIATE SOURCE:
;; LIBRARY: Adrenal
;; CLONE: 100877

US-09-193-524-1

Alignment Scores:

Pred. No.: 1-51e-80 Length: 1855
Score: 744.00 Matches: 186
Percent Similarity: 54.63% Conservative: 70
Best Local Similarity: 38.43% Mismatches: 128
Query Match: 30.32% Indels: 68
DB: 3 Gaps: 14

US-09-934-066-2 (1-466) x US-09-193-524-1 (1-1855)

QY 8 PheGlnIleValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
DB 60 TGGAAAGTAGTGTCTCTCAGTGTGGCCCTGGGAATTGGTGCC----- 104
QY 28 GlnLeuLeuAsnAspAspValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
DB 105 -----GTCTATAGATGATCCTGAAGATGGAGGCAAGCACTGG 143
QY 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
DB 144 GTGGTGATCGTGGCAGGTTCAATGCTGGTATATATAGGCACCGACGACCGCTGC 203
QY 68 HisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMet 87
DB 204 CATGCTACCACTTCAATCCCGCAATGGATCTCTGCGCAACAGATCGTTGTGATTATG 263
QY 88 TyrAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
DB 264 TACGATGACATAGCTTACTCTGAAGACAATCCCATCCAGGAATTGTGATCAACAGGCC 323
QY 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
DB 324 ANTGGCACAGATGCTATCAGGAGTCCCGAGGACTACACTGGAGAGGATGTTACCCCA 383
QY 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
DB 384 CAAATTTCTCTGTGTGTGAGAGCGGATGAGAGAGAGTGAAGGGTATAGGATCCGCC 443
QY 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
DB 444 AAGGCTCGAAGAGTGGTCCCGAGGATCAGCTGTTCTATTATTTCTACCATCATGATC 503
QY 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
DB 504 TCTGGNACTAGTGTTCCTCC-----AATGAAGATCTCATGTAAGAGCCTGATTAGACC 560

QY 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
DB 561 ACCATTACATTTTCAAAACAAATGTTACCGAAGATGGTGTCTACATTAGGCGCTGT 620
QY 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
DB 621 GAGTCTGGGTCCATGATGAACAC-----CTGCGGATAAATCAATGTTTATGCAACTACT 677
QY 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
DB 678 GCTGCCAACCCAGAGAGTCTCTACGCTGTACTATCATCAG----- 722
QY 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
DB 723 -----AAGAGGTCCACGTACCTGGGGACTGTACAGCGTCAACTGGATGGAAGACTCG 776
QY 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArg 286
DB 777 GACGTGGAAGATCTGACTAAGAGACCCGACACAGCATACCCCTGGTAAATCGCAC 836
QY 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
DB 837 ACC-----AACACCACCGCAGTCATCGACGTATGGAACAAACAATCTCC 881
QY 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
DB 882 ACCATGAAGTGTGATGAGTTCAGGT-----ATGAAA 914
QY 327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
DB 915 CGCAAAGCCAGTCTCCCTCCCTCCCTACCTCCAGTCACACACTTGACCTCAGCCCGCAGC 974
QY 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
DB 975 CCTGATGTCCTCTCACCATCATGAAAAGAAAGATGATCAACACC-----AATGATCTG 1028
QY 362 GluSerLys-----GluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
DB 1029 GAGGAGTCCAGCAGCTCAGGAGGAGATCCAGCGGTATCTGGATGCCAGGCAC----- 1082
QY 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
DB 1083 -----CTCATCCGAGGTGAGGTGGAGCAG----- 1106
QY 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTrpAspCys 417
DB 1107 -----CTCTCTCCGAGAGGCC-----CCGCTCAGCGGGCAC---AGCTGC 1145
QY 418 PheLysThrLeuValAsnSerPheLysAsnHisCys 429
DB 1146 TACCCAGAGGTCTGTGTGTACTTCCGAGCCACTGCG 1181

RESULT 5

US-09-928-613-15
; Sequence 15, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

1  OPERATING SYSTEM:  PC-DOS/MS-DOS
2  SOFTWARE:  PatentIn Release #1.0, Version #1.30
3  CURRENT APPLICATION DATA:
4  APPLICATION NUMBER:  US/08/928,613
5  FILING DATE:
6  CLASSIFICATION:  524
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER:  08/567,506
9  FILING DATE:
10 ATTORNEY/AGENT INFORMATION:
11 NAME:  Luther, Barbara J.
12 REGISTRATION NUMBER:  33954
13 REFERENCE/DOCKET NUMBER:  PP-0048 US
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE:  415-855-0555
16 TELEFAX:  415-852-0195
17 INFORMATION FOR SEQ ID NO:  15:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH:  220 base pairs
20 TYPE:  nucleic acid
21 STRANDEDNESS:  single
22 TOPOLOGY:  linear
23 MOLECULE TYPE:  cdna
24 IMMEDIATE SOURCE:
25 LIBRARY:  kidney
26 CLONE:  158868
27 15-08-928-613-15

```

Alignment Scores:		
Pred. NO.:	4,88e-13	Length:
Score:	178.00	Matches:
Percent Similarity:	62.20%	Conservative:
Best Local Similarity:	45.12%	Mismatches:
Query Match:	7.25%	Indels:
DB:	2	Gaps:
		220
		37
		14
		19
		12
		2

US-09-934-066-2 (1-466) x WS-08-928-613-15 (1-220)

216	LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp	235
QY		
DB		
10	CTGGCGATACATCAATGTTATGCAACCTGTCGCCAACCCCAAGAGATCGTCTCAT	69
QY		
236	GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly	255
QY		
DB		
70	GCCTGTAACTATGATGAG-----AAGAGGTCCACGTAACCTGGGG	108
QY		
256	AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr	275
QY		
DB		
109	GACTGGTACAGCGTCACTCGATGGAGAGACTCGAGCTGGAGAACTGACTAAAGAGACC	168
QY		
276	LeuGluGlnGlnTyrHisValValLysArgArgValGlySerAspValProGluThrSer	295
QY		
DB		
169	CTGCACAGCAGTACCACCTGGTAAAAATGCGACACC-----AACACCACGC	213
QY		
296	HisVal	297
QY		
214	CACGTT	219
DB		

RESULT 6

US-09-193-324-15
Sequence 15, Application US/09193524
Patent No. 6007997
GENERAL INFORMATION:
APPLICANT: Diep, Dinh
APPLICANT: Braxton, Scott M.
APPLICANT: Deleagane, Angelo M.
TITLE OF INVENTION: A NOVEL HUMAN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICAL
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,524
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,613
FILING DATE:
APPLICATION NUMBER: 08/567,506
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PP-0048 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
IMMEDIATE SOURCE:
LIBRARY: kidney
CLONE: 158868
US-09-193-524-15

Alignment Scores:		
Pred. No.:	4,888-13	Length:
Score:	178.00	Matches:
Percent Similarity:	62.20%	Conservative:
Best Local Similarity:	45.13%	Mismatches:
Query Match:	7.25%	Indels:
DB:	3	Gaps:
		220
		37

US-09-934-066-2 (1-466) x US-09-193-524-15 (1-220)

216	LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp	235
QY		
Db		
10	CTGCGGGATAACCAATGTTATGCAACTACTGCTGCCAACCCACGAGAGTCGTCCTAC	69
QY		
236	GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly	255
QY		
Db		
70	GCCTGTACTATGATGAG-----AAGAGGTCCACGTACCTGGCG	108
QY		
256	AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr	275
QY		
Db		
109	GACTGGTACACGCGTCAACTGGATGGAGAGCTCGGACCTGGAAACACTGTACTAAAGAGAC	168
QY		
276	LeuGluGlnGlnTyrHisValValLysArgValGlySerAspValProGluThrSer	295
QY		
Db		
169	CTGCACAAGCAGTACCACCTGGTAAATATCGCACACC-----AACACCACGC	213
QY		
296	HisVal	297
QY		
Db		
214	CACGTT	219
Db		

RESULT 7

US-08-928-613-11
; Sequence 11, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleagane, Angelo M.

;; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE

;; NUMBER OF SEQUENCES: 24

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

;; STREET: 3174 Porter Drive

;; CITY: Palo Alto

;; STATE: CA

;; COUNTRY: USA

;; ZIP: 94304

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/928,613

;; FILING DATE:

;; CLASSIFICATION: 524

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/567,506

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Luther, Barbara J.

;; REGISTRATION NUMBER: 33954

;; REFERENCE/DOCKET NUMBER: PF-0048 US

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-855-0555

;; TELEFAX: 415-852-0195

;; INFORMATION FOR SEQ ID NO: 11:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 219 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: cDNA

;; IMMEDIATE SOURCE:

;; LIBRARY: Bone Marrow

;; CLONE: 104286

;; US-08-928-613-11

Alignment Scores:

Pred. No.: 8.04e-12 Length: 219

Score: 168.00 Matches: 34

Percent Similarity: 60.98% Conservative: 16

Best Local Similarity: 41.46% Mismatches: 20

Query Match: 6.85% Indels: 12

DB: 2 Gaps: 2

US-09-934-066-2 (1-466) x US-08-928-613-11 (1-219)

Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241

Db 1 GTTATGCAACTNCTGCTGCCAACCCAGAGAGTCGTCCTACGGCTGTNACTATGATGAG 60

Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261

Db 61 -----AAGAGGTCACCGTACCTGGGGGACTGGTACAGCGTCAAC 99

Qy 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHis 281

Db 100 TGGATGGAAGACTCGGACGTTGGAAGATCTGCTACTAAAGAGACCTCGACAGAGTACCAC 159

Qy 282 ValValLysArgValGlySerAspValProGluThrSerHisValCysArgPheGly 301

Db 160 CTGTAAATCGCACACC-----AACACCAGCCAGCGTCATGCACTATGGA 204

Qy 302 ThrGlu 303

Db 205 AACAAA 210

RESULT 8

US-09-193-524-11

; Sequence 11, Application US/09193524

;; Patent No. 6007997

;; GENERAL INFORMATION:

;; APPLICANT: Diep, Dinh

;; APPLICANT: Braxton, Scott M.

;; APPLICANT: Delegeane, Angelo M.

;; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE

;; NUMBER OF SEQUENCES: 24

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

;; STREET: 3174 Porter Drive

;; CITY: Palo Alto

;; STATE: CA

;; COUNTRY: USA

;; ZIP: 94304

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/193,524

;; FILING DATE:

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/928,613

;; FILING DATE:

;; APPLICATION NUMBER: 08/567,506

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Luther, Barbara J.

;; REGISTRATION NUMBER: 33954

;; REFERENCE/DOCKET NUMBER: PF-0048 US

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-855-0555

;; TELEFAX: 415-852-0195

;; INFORMATION FOR SEQ ID NO: 11:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 219 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: cDNA

;; IMMEDIATE SOURCE:

;; LIBRARY: Bone Marrow

;; CLONE: 104286

;; US-09-193-524-11

Alignment Scores:

Pred. No.: 8.04e-12 Length: 219

Score: 168.00 Matches: 34

Percent Similarity: 60.98% Conservative: 16

Best Local Similarity: 41.46% Mismatches: 20

Query Match: 6.85% Indels: 12

DB: 2 Gaps: 2

US-09-934-066-2 (1-466) x US-09-193-524-11 (1-219)

Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241

Db 1 GTTATGCAACTNCTGCTGCCAACCCAGAGAGTCGTCCTACGGCTGTNACTATGATGAG 60

Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261

Db 61 -----AAGAGGTCACCGTACCTGGGGGACTGGTACAGCGTCAAC 99

Qy 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHis 281

Db 100 TGGATGGAAGACTCGGACGTTGGAAGATCTGCTACTAAAGAGACCTCGACAGAGTACCAC 159

Qy 282 ValValLysArgValGlySerAspValProGluThrSerHisValCysArgPheGly 301

Db 160 CTGTAAATCGCACACC-----AACACCAGCCAGCGTCATGCACTATGGA 204

Qy 302 ThrGlu 303
Db 205 AACAAA 210

RESULT 9
US-09-134-001C-1411
: Sequence 1411, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1411
: LENGTH: 1860
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1411

Alignment Scores:
Pred. No.: 1,78e-06 Length: 1860
Score: 137.00 Matches: 104
Percent Similarity: 36.57% Conservative: 73
Best Local Similarity: 21.49% Mismatches: 191
Query Match: 5.58% Indels: 116
Gaps: 22

US-09-934-066-2 (1-466) x US-09-134-001C-1411 (1-1860)

Qy 22 AlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLysSer 41
Db 208 GCTGAATACGTAAGTGGTGCAGATAATAGTATGATGAGGATGGTACGTTAGGA 258
Qy 42 AlaLysGlyThrArgTrpAlaValLeuAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 259 ATTGGTCATACAGCTGGGCG-----ACACATGGTGTGCCAAATATGAAATTCACAT 312
Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp--- 80
Db 313 CCACACAGTCACATCTGCAGCTTTTACATATTAGTTCATATGATGGTAAATGAAATATAT 372
Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 373 GAGAATTAAGCTGAATATTATCTGATGTCATCTTTTTCATCAGAACT----- 423
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 424 -----GATACGGAAGTTATTGTACAAATAGTA-----GATTAT 456
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuGlyAsnGluSerGly 140
Db 457 -----TTTTCTAGACAAGGATTAGCTACAGAAGATGCA 489
Qy 141 ValThrGlyGlyAsnGlyLysValValLys----- 150
Db 490 TTTTACA-----AAAGTAGTTAAATATTATCATGGTGCATATGCTTTAGGATTA 537
Qy 151 SerGlyProAsnAspAsnIlePheIleTyrTrpAlaAspHisGlyAlaProGlyLeuIle 170
Db 538 TTAGATGATAATGATAAGATACTATTTATGCTGGCTAAAAACAAGCTCTCCGCTTTAGTA 597
Qy 171 AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuLysMet 190
Db 598 GGTGTAGGTGAAGTTTCAATGTTTCTGATGCTAGCAATGCTTACAACTACA 657
Qy 191 HisLysArgLysLys-----TyrAsnLysMetValIleTyrValGluAlaCysGlu 207

Db 658 AACCAATACAAAGAGATACATCACCATGAATAGTTATTGTTAAGCGAGACACAGTAGAA 717
Qy 208 SerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAla 227
Db 718 ATTAAAGATCTTGAGGGGCACATTCACACACGTGATACGTATACGGCAGAAATAGATGCT 777
Qy 228 AlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro 247
Db 778 GCTGATGCAGAAAAGCGGTATATGATCATTTAAAGAAATTCATGAACAGCCT 837
Qy 248 Ser-----GluIleGlyThrCysLeuGlyAsp 256
Db 838 GCAGTGATCGTCGCATTATTCAAGAATATCAAGATGAAAAGGTAAATTTAAAATCGAT 897
Qy 257 ThrPheSerIleSerTrpLeuGluAspSerAsp---LeuHisAspMetSerLysGluThr 275
Db 898 TCAGAGATTATTAAATGATGTAGCAGATGCTGATATTTACATCTGTCAGCTGCTACT 957
Qy 276 LeuGluGlnGlnThrHisVal-----ValLysArgArgValGly 288
Db 958 -----AGTTATCATCTCGATTGGTTGCTAAAGAAATTTATTGAAAATGGCAGGT 1008
Qy 289 SerAspValProGluThrSerHisValCysArg-----PheGlyThrGluLysMet 305
Db 1009 -----GTACCTACTCAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1062
Qy 306 LeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu----- 319
Db 1063 TCTGAAAACCACTATTTATTTATTTTACCAANTCTGTGAAACAGCTAGTAGCTGCTGCT 1122
Qy 320 -----AsnAspAsnPheThrPheThrGluSerPheSer 330
Db 1123 GTATTAGTTGAAACAAATAAGTTAGTTCACAAATCATTAACAATTTACTAATGTTGCTGCT 1182
Qy 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
Db 1183 TCAACATTTATCAGCTGAAGCG-----GATCATCATATTACTTTTACATGCTGCTGGA 1230
Qy 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
Db 1231 CCGTACATTCGATCGCATCTACAAAAGCATATACAGCGCAAAATGCTGTTTATCTATCT 1290
Qy 371 LeuAspGlu-----LysAsnHisArgLysGlnIleAspGlnSerIleThrAspIle 387
Db 1291 TTATCTCAAATGTTGCTTAAATCATGCTGCTGGAACCGAT-----CTTGATTATA 1341
Qy 388 LeuArgLeuSerValLysGlnThrAsnVal----- 407
Db 1342 TTAAAGAACTAGCTAAGGTTACTACAGCTATT----- 1374
Qy 408 GlyGlnProLeuValAspAsp-----TrpAspCysPheLys 419
Db 1375 ---GAAACAATTTGTCAGCATGCTACCTAAGATGGAGCAATTTGCACCGGATTTCTTAA 1431
Qy 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 1432 ACTACTCGTAATGCATTC-----TTTATTGGACCAACAATTTGATTATAATGTTAGTTA 1485
Qy 440 ThrGlyAlaLeu 443
Db 1486 GAAGGTGCATTA 1497

RESULT 10
US-09-280-116-247
: Sequence 247, Application US/09280116A
: Patent No. 6331427
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith F.
: TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
: FILE REFERENCE: 5800-24, 035800/176965
: CURRENT APPLICATION NUMBER: US/09/280,116A
: CURRENT FILING DATE: 1999-03-26

```
;
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: hemoglobinase
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(544)
; OTHER INFORMATION: n = a, t, c or g
;
; US-09-280-116-247
;
; Alignment Scores:
; Pred. No.: 2.98e-07 Length: 544
; Score: 136.00 Matches: 25
; Percent Similarity: 70.45% Conservatives: 6
; Best Local Similarity: 56.82% Mismatches: 13
; Query Match: 5.54% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-934-066-2 (1-466) x US-09-280-116-247 (1-544)
;
; QY 53 GlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIle 72
; DB 34 GGATCTAAATCTACCAATCATTTCTCCAGGACGACGCGTCCATCGCTACCAATC 93
;
; QY 73 LeuArgLysGlyGlyLeuLysaspGluAsnIleIleValPheMetTyrAspIleAla 92
; DB 94 ATTCAACCGCAATGGGATTCCTCAGCAAGACATCGTTGTGATGATGATGATGATGAT 153
;
; QY 93 PheSerSerGlu 96
; DB 154 TACTCTGAAGAG 165
;
; RESULT 11
; US-09-928-613-5
; Sequence 5, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/928,613
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: THP-1
; CLONE: 075848
;
; US-09-928-613-5
;
; Alignment Scores:
; Pred. No.: 4.15e-08 Length: 155
; Score: 135.50 Matches: 25
; Percent Similarity: 69.23% Conservatives: 11
; Best Local Similarity: 48.08% Mismatches: 11
; Query Match: 5.52% Indels: 5
; DB: 2 Gaps: 1
;
; US-09-934-066-2 (1-466) x US-09-928-613-5 (1-155)
;
; QY 252 ThrCysLeuGlyAspThrPheSerIleSerTyrLeuGluAspSerAspLeuHisAspMet 271
; DB 13 AGTACTCTGGGACCTGGTACAGCTCACTCGATGGAGACCTCGGACCTGGAGATCTG 72
;
; QY 272 SerLysGluThrLeuGluGlnGlnTyrHisValValLysArgValGlySerAspVal 291
; DB 73 ACTAAGACACCTGCACAGCAGTACCACCTGGTAAATCGCACACC----- 120
;
; QY 292 ProGluThrSerHisValCysArgPheGlyThrGlu 303
; DB 121 ---AACACACGACGCTCATGTCAGTATGGAACAAA 153
;
; RESULT 12
; US-09-193-524-5
; Sequence 5, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/193,524
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,613
; FILING DATE:
; APPLICATION NUMBER: 08/567,506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: THP-1
; CLONE: 075848
; US-09-934-066-2 (1-466) x US-09-934-066-2 (1-466)
;
Alignment Scores:
Pred. No.: 4.15e-08 Length: 155
Score: 135.50 Matches: 25
Percent Similarity: 69.23% Conservatives: 11
Best Local Similarity: 48.08% Mismatches: 11
Query Match: 5.52% Indels: 5
DB: 3 Gaps: 1
;
US-09-934-066-2 (1-466) x US-09-934-066-2 (1-466)
;
; 252 ThrCysLeuGluAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMet 271
; 13 ACGTACCTGGGGGACTGCTGACAGCTCACTGAGTGGGAGACTCGGAGATCTG 72
;
; 272 SerLysGluThrLeuGluGlnGlyHisValValLysArgValGlySerAspVal 291
; 73 ACTAAGAGACCTCCACCAAGCAGTACCCTGTTAAATGCCACACC----- 120
;
; 292 ProGluThrSerHisValCysArgPheGlyThrGlu 303
; 121 ---ACACGACGACCTGATGATGATGGAACAAA 153
;
RESULT 13
US-09-934-066-2 (1-466) x US-09-934-066-2 (1-466)
; Sequence 13, Application US/08928613
; Patent No. 5840562
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleage, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928.613
; FILING DATE:
; CLASSIFICATION: 524
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/567.506
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 13:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Lung
; CLONE: 125569
; US-09-928-613-13
;
Alignment Scores:
Pred. No.: 1.92e-05 Length: 229
Score: 116.00 Matches: 31
Percent Similarity: 56.76% Conservatives: 11
Best Local Similarity: 41.89% Mismatches: 22
Query Match: 4.73% Indels: 10
DB: 2 Gaps: 2
;
US-09-934-066-2 (1-466) x US-09-928-613-13 (1-229)
;
; 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
; 5 TACATTGGAGCTCTGAGTCTGGGTCCATGATGAACCAC---CTNCCGGGATAACATCAAT 61
;
; 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
; 62 GTTATGCAACTACTGCTGCCAACCCAGAGAGTCTCTACCCCTGTACTATGATGAG 121
;
; 242 SerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle-Se 261
; 122 -----AAGAGNCCAGCTACCTGGGGGACCTNGTACAAANGTAA 160
;
; 261 rTrpLeuGluAsp-SerAspLeuHisAspMetSerLys 273
; 161 NTNGATGGAAGATTCACAGCAGGAGGAGATCTNNCTAAA 198
;
RESULT 14
US-09-934-066-2 (1-466) x US-09-928-613-13 (1-229)
; Sequence 13, Application US/09193524
; Patent No. 6007997
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; APPLICANT: Braxton, Scott M.
; APPLICANT: Deleage, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/193.524
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928.613
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0048 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 13:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

LENGTH: 229 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
LIBRARY: Lung
CLONE: 125569
US-09-193-524-13

Alignment Scores:
Pred. No.: 1.92e-05 Length: 229
Score: 116.00 Matches: 31
Percent Similarity: 56.76% Conservative: 11
Best Local Similarity: 41.89% Mismatches: 22
Query Match: 4.73% Indels: 10
DB: 3 Gaps: 2

US-09-934-066-2 (1-466) x US-09-193-524-13 (1-229)

Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsh 221
Db 5 TACATTGANGCCTGTGAGTCTGGTCCATGATGAACAC---CTNCCGGATAACATCAAT 61
Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyrCysProGlu 241
Db 62 GTTTATGCAACTACTGCTGCCAACCCAGAGTCGCTACGCGCTGTACTATGATGAG 121
Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle-Se 261
Db 122 -----AAGAGNCCACGTACCTGGGGGACTNGTACAAAGTNAAA 160
Qy 261 rTrpLeuGluAsp-SerAspLeuHisAspMetSerLys 273
Db 161 NTNGATGAGAAATTCAGACGAGGAGATCTNNCTAAA 198

RESULT 15

US-09-453-702B-256
Sequence 256, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Buriland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453.702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 256:
SEQUENCE CHARACTERISTICS:
LENGTH: 9510 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 256:
US-09-453-702B-256

Alignment Scores:
Pred. No.: 0.0116 Length: 9510
Score: 115.50 Matches: 115
Percent Similarity: 36.46% Conservative: 83
Best Local Similarity: 21.18% Mismatches: 188
Query Match: 4.71% Indels: 157
DB: 4 Gaps: 30

US-09-934-066-2 (1-466) x US-09-453-702B-256 (1-9510)

Qy 18 LeuIlePheSerAlaGluSerArgLysThrClnLeuLeuAsnAspAsn----- 33
Db 7486 ATTATTATTGTAATCATGGGCAAAATACACCTTCCTCAATTCAGGAAGTAAACATATA 7545
Qy 34 AspValGluSerSerAspLysSerAlaLys-----GlyThrArgTrpAlaVal 49
Db 7546 GATATTACACTCTTGTATAAAATAGTCATTATCATACTTTTGACTTCCTGAATATTATT 7605
Qy 50 LeuValAlaGlySerAsnGlu---TyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db 7606 TTTGAACATTACACCAATGAGATATTTATAAGTTTGTATAATCAAGGAGGATTT----- 7659
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db 7660 GTTATATCAATATTAAATAATGTACGAGTGAAGCGGCAACATTAATGATTTCGGAATA 7719
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db 7720 AACATGACTTCTTTAGATAGTAGT-----GGTTCATTGATATATTTCCCTAGC 7767
Qy 109 GlyGlu-----AspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 7768 GGTGATATTTATCATATTTTCAGATATCTATAAA---ATGCCAGAGGTAGAAAAATCGTTT 7824
Qy 124 AlaValAsnValGln-----AsnPheTyrAsnValLeuLeuGlyAsn 137
Db 7825 AAATTAATGTAGAAAGAAAGAACCTGACATTGATGATATCATCAATGTGGCTATTTTGA 7884
Qy 138 GluSerGlyValThrGlyGlyAsnGlyLysValLysSerGlyProAsn-----Asp 155
Db 7885 ACTCTTATCTACAA-----ATAAAAAAATCCCAATTAATGATGAT 7926
Qy 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
Db 7927 AGTGATTATATTTTATGTTTAGACAAC-----CCTAACCTCTCATCT----- 7968
Qy 176 AspGluValMetAlaLysAspPheAsnGluVal----- 186
Db 7969 -----TACACATTGAACCTTTAATGACTGTCTGGATACATATCAAGTTTATGGAT 8019
Qy 187 -----LeuGluLysMetHisLys-----Arg 193
Db 8020 AATATTAGGGGTAGTTTTTACCCTTTTCATAAAAAATACTGTAAACATCGCCCTTAATGAA 8079
Qy 194 LysLysTyrAsnLysMetValIleTyrValIcAlaCysGluSerGlySerMetPheGlu 213
Db 8080 AAAAAATATATTTCCCTTAATCGGACTTTGATAAGTTGTCAATTAATGATGTTTGTAGA 8139
Qy 214 GlyIleLeuLys---LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGlu 232
Db 8140 CAAGCATTAGAGGTAAAAAATAAGATAGTTACAGATAAGT-----AAA 8184

```
Qy 233 SerSerTyrGlyValTyr-----CysProGlu----- 241
Db : : : : :
8185 TTTACATGGGAACGATATCGTGTGATACCTGACCTGAAGATAGAAATTTCTCATCTG 8244
Qy 242 -----SerTyrProProProSerGluIleGlyThrCysLeuGlyAsp 256
Db : : : : :
8245 GAGTTAGATGGGTTTAATATTATTTTACAAACCC---GAGTTAGATACTCCAATATCTGAC 8301
Qy 257 ThrPheSerIleSerTyr-----LeuGluAspSerAspLeuHis----- 269
Db : : : : :
8302 TCATTTTCATATTTATATATTTTCAGATCGTTGATGATGTCATATTAAAGCTT 8361
Qy 270 ---AspMetSerLysGluThrLeuGluGlnTyrHis-----ValValLysArgArg 286
Db : : : : :
8362 CTTCACTCAACAGAGAAACAAAGCAATAACACACATCGTATTATTCTTTAAAGATAC 8421
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db : : : : :
8422 TTTATAGACTATTGCTTAAACAAGTATCACTGACACA-----GAAACAATATA 8472
Qy 307 -----LysAspTyrLeuSerSerTyrIleGlyArgAsn----- 317
Db : : : : :
8473 TATCCTGTGATTTGTCATTCACCTGATCCTTCACAGCGACATATATAGGCATCCATTT 8532
Qy 318 -----ProGluAsnAspAsnPheThrPheThrGlu 327
Db : : : : :
8533 AGAATAGTTCTCGGAATAAGACATTTGATCCATCCAGGAAGACTGTGAAAATTCATAAGC 8592
Qy 328 SerPheSerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeu 347
Db : : : : :
8593 ACCTCAAAAGAATATTAAAGCAATATGGATGTGATTAACAATGTTATTGTTCCCT----- 8646
Qy 348 TyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGln 367
Db : : : : :
8647 -----CAA 8649
Qy 368 LysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIle 387
Db : : : : :
8650 AAAACAACAAGAAACAAATATATCCATAGTTAGTTTAAATAGCAATATCAAAAATGAT 8709
Qy 388 LeuArgLeuSerValLysGlnThrAsnVal-----LeuAsn----- 399
Db : : : : :
8710 ATGTATTCTCGGAGTATGACCGGAACATCAAGATTTTTCATCTAAATAATTTCAGGT 8769
Qy 400 -----LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCys 417
Db : : : : :
8770 GATTATTCTTAACAACCTCAAAACTCATGGTGGTGGTGTG-----GTTATT 8820
Qy 418 PheLysThrLeuValAsnSerPhe-----LysAsnHisCysGlyAlaThrValHistyr 435
Db : : : : :
8821 TTAAAGATTTTATAAATACTGGTGGAAATATAATCTTACTTTAATAACAGTTCCTATT 8880
Qy 436 GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln 455
Db : : : : :
8881 GATAATAAATGTCAGACAAATAAATAATACCCCAATGGGGATAAAGATACAGGAA 8940
Qy 456 ThrValSer 458
Db : : : : :
8941 ACAGTCAGT 8949
```

Search completed: May 29, 2003, 19:47:09
Job time : 99 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 19:19:39 ; Search time 186 Seconds
(without alignments)
3378.162 Million cell updates/sec

Title: US-09-934-066-2
Perfect score: 2454
Sequence: 1 MSSPLGHQILVFLHALLIF.....CNMGVDVKQTVSAIEQACSM 466

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09934066/runat_19052003_163702_24460/app_query.fasta_1.647
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09934066 -CGN_1_1_80 -runat_19052003_163702_24460
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAEXT=0.5 -FPGAPOP=6 -FPGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2454	100.0	1560	10	US-09-934-066-1 Sequence 1, Appli
2	2188	89.2	4320	10	US-09-934-066-3 Sequence 3, Appli
3	1246	50.8	1485	9	US-09-938-842A-1485 Sequence 1485, Ap
4	1160	47.3	1461	9	US-09-938-842A-927 Sequence 927, App

5	825	33.6	1974	9	US-10-208-408-18	Sequence 18, Appli
6	818	33.3	2002	12	US-10-002-600-10	Sequence 10, Appli
7	744	30.3	1855	9	US-09-967-796-1	Sequence 1, Appli
8	487	19.8	960	9	US-10-043-487-97	Sequence 97, Appli
9	450	18.3	405	10	US-09-878-574-4049	Sequence 4049, Ap
10	441	18.0	393	10	US-09-878-574-3641	Sequence 3641, Ap
11	412.5	16.8	392	10	US-09-878-574-3407	Sequence 3407, Ap
12	333	13.6	415	10	US-09-960-352-13205	Sequence 13205, A
13	333	13.6	422	10	US-09-960-352-11028	Sequence 11028, A
14	333	13.6	426	10	US-09-960-352-1596	Sequence 1596, Ap
15	307	12.5	407	10	US-09-960-352-5110	Sequence 5110, Ap
16	289	11.8	415	10	US-09-960-352-10118	Sequence 10118, A
17	245.5	10.0	282	10	US-09-878-574-7312	Sequence 7312, Ap
18	241	9.8	455	9	US-09-918-995-8935	Sequence 8935, Ap
19	226.5	9.2	375	10	US-09-878-574-2361	Sequence 2361, Ap
20	210	8.6	184	10	US-09-960-352-11538	Sequence 11538, A
21	190.5	7.8	459	9	US-09-918-995-11442	Sequence 11442, A
22	178	7.3	220	9	US-09-967-796-15	Sequence 15, Appli
23	178	7.3	367	10	US-09-878-574-1704	Sequence 1704, Ap
24	168	6.8	219	9	US-09-967-796-11	Sequence 11, Appli
25	157	6.4	454	10	US-09-878-574-4350	Sequence 4350, Ap
26	151	6.2	273	10	US-09-923-876-4813	Sequence 4813, Ap
27	146	5.9	15061	9	US-10-092-154-991	Sequence 991, App
28	146	5.9	15061	9	US-09-764-891-6914	Sequence 6914, Ap
29	146	5.9	15061	10	US-09-764-847-991	Sequence 991, App
30	144.5	5.9	442	10	US-09-920-300A-729	Sequence 729, App
31	144.5	5.9	442	12	US-10-033-528-729	Sequence 729, App
32	135.5	5.5	155	9	US-09-967-796-5	Sequence 5, Appli
33	133.5	5.4	280	10	US-09-878-574-5529	Sequence 5529, Ap
34	127	5.2	3309400	9	US-09-738-626-1	Sequence 1, Appli
35	116	4.7	229	9	US-09-967-796-13	Sequence 13, Appli
36	115.5	4.7	9510	9	US-10-114-170-256	Sequence 256, App
37	114.5	4.7	331	9	US-09-967-796-10	Sequence 10, Appli
38	114.5	4.7	1743	9	US-10-202-211-1	Sequence 1, Appli
39	108.5	4.4	1074	10	US-09-974-300-2794	Sequence 2794, Ap
40	107	4.4	2988	9	US-09-738-626-347	Sequence 347, App
41	104.5	4.3	206	9	US-09-967-796-22	Sequence 22, Appli
42	104.5	4.3	244	10	US-09-878-574-10299	Sequence 10299, A
43	104	4.2	4959	9	US-10-025-950-1	Sequence 1, Appli
44	103.5	4.2	199	9	US-09-967-796-18	Sequence 18, Appli
45	103.5	4.2	12379	10	US-09-991-258-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-09-934-066-1
; Sequence 1, Application US/09934066
; Patent No. US20020108149A1
; GENERAL INFORMATION:
; APPLICANT: Gruijs, Darren B.
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide Accumulation in Plants
; FILE REFERENCE: 35718/237251
; CURRENT APPLICATION NUMBER: US/09/934,066
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,804
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-934-066-1
Alignment Scores:
Pred. No.: 5.3e-279 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

DB	Accession	Length	Score	Percent Similarity	Best Local Similarity	Query Match	DB			
Db	1060	CCGCGCGCATATTCCTCTGCTATACCTCCAGAGAAAGATTCCAAAAGCTCCAAATGGGATCA	1119							
Qy	361	LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHIsArGlyGlnIle	380							
Db	1120	CTTGAAGAGCAAGAGCTCAGAGAAATTCCTTGACGAAAAGAAATCATAGGAACAATC	1179							
Qy	381	AspGlnSerIleThrAspIleLeuArGLeuSerValLysGlnThrAsnValLeuAsnLeu	400							
Db	1180	GATCAGAGCATTTACAGACATTCCTGCGCTTTCAGTTAAACAAACCAATGCTTTAAATCTC	1239							
Qy	401	LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspThrAspCysPheLysThr	420							
Db	1240	TTAACTTCCACAAGAACACAGGACGCTCTTGTAGACGATTTGGGATTTGCTTCAAGACT	1299							
Qy	421	LeuValAsnSerPheLysAsnHisCysGlyValaThrValHisThrGlyLeuLysThr	440							
Db	1300	CTAGTTAATAGCTTCAGAAATCACTGCGGTGCACCGGTGCATTTACGGATTTGAATATACA	1359							
Qy	441	GlyAlaLeuAlaAsnIleCysAsnMetGlyValaAspValLysGlnThrValSerAlaIle	460							
Db	1360	GGAGCGCTGCCAATATCTGCAATATGCGAGTGGATGGATGAAGCAAACTGTTTCAGGCATTT	1419							
Qy	461	GluGlnAlaCysSerMet 466								
Db	1420	GAACAAGCTTTGTTGCGATG 1437								
RESULT 2										
US-09-934-066-3										
; Sequence 3, Application US/09934066										
; Patent No. US20020108149A1										
; GENERAL INFORMATION:										
; APPLICANT: Gruis, Darren B.										
; APPLICANT: Jung, Rudolf										
; TITLE OF INVENTION: Methods of Increasing Polypeptide										
; FILE REFERENCE: 35718/237251										
; CURRENT APPLICATION NUMBER: US/09/934.066										
; CURRENT FILING DATE: 2001-08-21										
; PRIOR APPLICATION NUMBER: US 60/226,804										
; PRIOR FILING DATE: 2000-08-21										
; NUMBER OF SEQ ID NOS: 3										
; SOFTWARE: FastSeq for Windows version 4.0										
; SEQ ID NO 3										
; LENGTH: 4320										
; TYPE: DNA										
; ORGANISM: Arabidopsis thaliana										
US-09-934-066-3										
Alignment Scores:										
Pred. No.:	5	91e-247	Length:	4320						
Score:	2188.00	Matches:	463							
Percent Similarity:	67.89%	Conservative:	0							
Best Local Similarity:	67.89%	Mismatches:	3							
Query Match:	89.16%	Indels:	219							
DB:	10	Gaps:	7							
US-09-934-066-2 (1-466) x US-09-934-066-3 (1-4320)										
Qy	1	MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20							
Db	2020	ATGTCTAGTCCCTCTGGTCACTTTTCAGATCTCTGTTTCTTCATGCTTTGCTTATCTTC	2079							
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnAspValGluSerSerAspLys	40							
Db	2080	TCAGCTGAGTCCCGCAAAACCCCAATTTGCTCAACGATAATGATGTTCAATCTACGCCACAAG	2139							
Qy	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTrpTyrAsnTyr	60							
Db	2140	AGTGCAAAAGGCACACGATGGGCTGTTTAGTTGCTGGAT								

Db 2200 AGGCATCAGGTGTTAAATTAAGTTTGAACGTTTAAACATAACAACAAAAAGGTCCA 2259
Qy 63 -----GlnAlaAspIleCysHis 68
Db 2260 AGCGAGATTTGTATGAACATAACACCGAGGTTTATTTTACAGAGGCTGACATATGCCAC 2319
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysaspGluAsnIleIleValPheMetTyr 88
Db 2320 CGGTATCAGATACCTCCGAAAGGCGGTTTAAAGATGAAACATCATTTGTGTATGTAT 2379
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db 2380 GATGATATCGCGTTTCTCGAGAACTTCTAGGCTGAGGTATCATTAATAAACACAGAT 2439
Qy 109 GlyGluAspValTyrLysGlyValProLys ----- 118
Db 2440 GGAGAAGATGTTTATAAGGAGTTTCCCTAA - GGTTCTTATTTCTACTTCTTTTGTGCGTTA 2498
Qy 118 ----- 118
Db 2499 TTTCTAGCTTGAATTCATATATATATATCAAGTTTGTGTTATTTGGTGGTAG 2558
Qy 119 AspTyrThrLysGluAlaValAsnValClnAsnPhetYrAsnValLeuLeuGlyAsnGlu 138
Db 2559 GACTACACTAAAGAGCTGTTAATGTTCAAACCTTCTACAAATGTGTACTTTGGAAATGAA 2618
Qy 139 SerGlyValThrGlyLysGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe 158
Db 2619 AGTGGCTFCACAGCAGGAAATGGCAAGTTGTGAAAGTGTCTCTAATGATAATATCTTC 2678
Qy 159 IleTyrThrAlaAspHisGlyAlaProGlyLeuIle ----- 170
Db 2679 ATCTATTATGCTGACCATGGAGCTCTCGCTTAATAGGTTTCTTAATTTTATGAAATTA 2738
Qy 170 ----- 170
Db 2739 TTAGCTACCATCAATCCATATCTATAATAAAGATTTTCTCTTGATCTAGCAACCGCG 2798
Qy 171 -----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu 187
Db 2799 ATTTTCTCAGCGATGCCACTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGTCTTG 2858
Qy 188 GluLysMetHisLysArgLysLysTyrAsnLysMetVal ----- 200
Db 2859 GAGAAGATGCATAAGAGAAAAAATACAACAAGATGGT - ATATAACTCAACCATTCGTGTTA 2917
Qy 200 ----- 200
Db 2918 CCTAGCTTTATACATATGTGTTCTGTTTGAATCTCTATGGTGTGTTTTTGGATGTT 2977
Qy 201 -----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218
Db 2978 TAGGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAA 3037
Qy 219 AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr 238
Db 3038 AATCTCAACATATACGAGTGACTGCTGCTAATTTCTAAAGACAGCAGTGGGGAGTTTAC 3097
Qy 239 CysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPhe 258
Db 3098 TGTCTCAGCATATATCT 3157
Qy 259 SerIleSerTrpLeuGlu ----- 264
Db 3158 AGCATCTCTTGGCTTGAGGACAGCTACTGCAACAAAAAGATTCATCTTATGGACTA 3217
Qy 264 ----- 264
Db 3218 TTGCAATGATTTGTTTCTTGAAGAATATTTGTTCAATTTGTTCTATGTTTGTGTGT 3277
Qy 265 -----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisVa 282
Db 3278 GTTTGGGACAGTGCCTTCATGACATGAGCAAGAGACTTTGGAGCAACAATACCAGT 3337

Qy 282 lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyTh 302
Db 3338 TGTAAAGAGAAGACTAGGATCTGATGTACCAGACATCTCTCATGTATGCCGTTTCGGAAC 3397
Qy 302 rGluLysMetLeuLysaspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAs 322
Db 3398 AGAGAAGATGCTTAAAGATTTATCTTTCTCTTACATTTGGAAGAAATCTCGAAAAACGATAA 3457
Qy 322 nPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsnProAr 342
Db 3458 CTTTCACITTTACGGAATCTTTTCTCTCACCNAATCTCTAATCTCGCTTGTGTCATATCCGCG 3517
Qy 342 gAspIleProLeuLeuTyrLeuGlnArgLysIle ----- 353
Db 3518 CGATATTCTCTGCTATACCTCCAGAGAAAGGT - GAGCTTTTTCGGGTTTTTTTTCATCAT 3576
Qy 354 -----GlnLysAl 356
Db 3577 TTTAAACGAAAGAGCTTTTTCAGCATGTTTAAATGTTTATTCATCTCTTAGATTTCAAAAAGC 3636
Qy 356 aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHi 376
Db 3637 TCCATGGGATCACTTGAAGCAAGAGAGCTCAGAGAAATTTGCTTGACGAAAAAGATCA 3696
Qy 376 sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs 396
Db 3697 TAGCAACAATTCATCAGACATTTACAGACATTTCTCGCGCTTTTCAGTTAAACAACCAAA 3756
Qy 396 nValLeuAsnLeuLeuThrSerThrArgThrGlyGlnProLeuValAspAspTrpAs 416
Db 3757 TGTCTTAAATCTCTTAACCTTCCACAAGAAACAACAGACAGCCTCTTGTAGACGATTGGGA 3816
Qy 416 pCysPheLysThrLeu ----- 421
Db 3817 TTGCTTCAAGACTCTAGTAACAACACATCTCAAACTTGTACTTGTGTTCTACGCAA 3876
Qy 422 -----ValAsnSerPh 425
Db 3877 CAACCATTTGCATTATTACTAAACCCAGTGTATATCGAATGAAATCGCAGGTTTAAATAGCTT 3936
Qy 425 eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs 445
Db 3937 CAAGAATCACTGCGGTGCACAGGTCATTCAGGATTGAAGTATACAGAGCGCTTGCCAA 3996
Qy 445 nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSe 465
Db 3997 TATCTGCAATATGGAGTGGATGTGAAGCAAACTGTTTTCAGCCCATTTGAACAGCTTTGTTT 4056
Qy 465 rMet 466
Db 4057 GATG 4060

RESULT 3

US-09-938-842A-1485
: Sequence 1485, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: FILE REFERENCE: SCRI1300-3
: CURRENT APPLICATION NUMBER: US/09/938, 842A
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264, 647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300, 111
: PRIOR FILING DATE: 2001-06-22


```
Qy 26 -----LysThrGlnLeuLeuAsnAspAsnValGluSer 37
Db 76 CGGTTCCAGGCAAAAGATTCTTATGCCGAGGAGGAGCAAGCAAGCGGTGAC----- 126
Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 127 CAAGCAGCAAGATGGTGCAGATGAGGCGGTTCTCGTGGTGGTCTTCTTGGGATAT 186
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77
Db 187 GGAACACTACAGACACAGGCTGAGTGTGCACGCATATCAAACTACTAAGAAAAGAGGT 246
Qy 78 LeuLysAspGluAsnIleIleValPheMetTyrAspIleAlaPheSerSerGluAsn 97
Db 247 TTAAGGAAGAAAACATAGTCGTTTGTATGATGATGATATCGCAAAACCCACTTAAT 306
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 307 CCTCGTCCGGTACTCTCAACCATCTGACGGTGACGATGTTTACGCCGGAGTCCCT 366
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 367 AAGGACTATAGTGTAGTACGTTTACGGCTGCAAACTTCTACGCTGTACTCTAGGCGAC 426
Qy 138 GluSerGlyValThrGlyLysGlyAsnGlyLysValLysSerGlyProAsnAspAsnIle 157
Db 427 CAGAAGGCTGTTAAGGTGAAGCGGTGAAGTCACTGCTAGCAAGCCCAACGATCACTT 486
Qy 158 PheIleTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 487 TTCGTATATTATGCGGATCATGGTGGTCCCGGAGTCTTGGGATGCGCAAAATACGCCCTCAC 546
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMethisLysLysLysTyrAsn 197
Db 547 ATATATCCAGCTGATTTTATTGAACGCTTAAGAAGAAGCATGCTCCGGACATACAAA 606
Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys 217
Db 607 GAGATGTTTATACGTAGAGCGTGTGAAGTGGGAGTATTTTCGAAGGGATAATGCCA 666
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 667 AAGGACTTGACATTTACGTAAACACGCGTTCAATGCCAAGAAGATGATTATGGAAACA 726
Qy 238 TyrCysProGluSerTyrProProSerGluIleGlyThrCysLeuGlyAspThr 257
Db 727 TATTGCTCTGGCATGAATCCGTCACCCCATCTGAAATATATCACTTCTTAGGGGATTTA 786
Qy 258 PheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGlu 277
Db 787 TATAGTCTTCTGGATGGAAGATAGTGAGACTCACAAATTTAAGAAAGAGACCATTAAG 846
Qy 278 GlnGlnTyrHisValValLysArgArgValGly-----SerAspValProGluThrSer 295
Db 847 CACAATACCACAGCGTGAAGATGAGGACATCAAACTACATACCTACTCAGTGGCTCT 906
Qy 296 HisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGly 315
Db 907 CATGTGATGAATACGGTACAATAGTATTAAAGTCGGAGAAGCTTTATCTTTACCAAGG 966
Qy 316 ArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSerProIle---Ser 334
Db 967 TTTGATCCACCCCGGTTAATCTCCCACTAAACGAA-----TTTACCGGTCAAGTCA 1017
Qy 335 AsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGln 354
Db 1018 AATAGAGATGCTTAAACCAACGGCAGCGGACCTTCTTCTTCTTGGCATATGATCGG 1077
Qy 355 LysAlaProMetGlySerLeuLeuSerLysGlyAlaGlnLysLysLeuLeuAspGluLys 374
Db 1078 ACATCGGAAGATGGTCAAGGAAGAGGATGACACATTTGAAGGAATTAATGAGACAACA 1137
Qy 375 AsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGln 394
```

```
Db 1138 AGCATAGGAACAATTTAGATGCAAGCGTCAAAATTTGATAGCCACAATTTTGTGGTCCG 1197
Qy 395 Thr---AsnValLeuAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 413
Db 1198 ACATGAAATGCTTAACTTAACTTGGTT-----AGAAACCCGGTTTGGCTTGGTTGAC 1248
Qy 414 AspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrVal 433
Db 1249 GATTGGCAATGCTTAAATCGATGCTGCTGTTTCAAGAGCATTTCTGGATCACTAACG 1308
Qy 434 HisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspVal 453
Db 1309 CAATATGGGATGAACATATCGGAGCGTTTGCACAACTTTGTAACAACGGTGTGTGCCAA 1368
Qy 454 LysGlnThrValSerAlaIleGluGlnAlaCys 464
Db 1369 GAGCTGATGAGGAGGAGGCTTCTACTCGCGCATGC 1401

RESULT 5
US-10-208-408-18
; Sequence 18, Application US/10208408
; Publication No. US20030096272A1
; GENERAL INFORMATION:
; APPLICANT: Scheybe, Xiao Min
; TITLE OF INVENTION: GENES REGULATED BY PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR
; FILE REFERENCE: PA-0048-1 US
; CURRENT APPLICATION NUMBER: US/10/208,408
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: 60/308,868
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PERL Program
; SEQ ID NO 18
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030096272A1 1301477CB1
US-10-208-408-18

Alignment Scores:
Pred. No.: 6.84e-87 Length: 1974
Score: 825.00 Matches: 179
Percent Similarity: 55.13% Conservative: 79
Best Local Similarity: 38.25% Mismatches: 158
Query Match: 33.62% Indels: 52
DB: Gaps: 13

US-09-934-066-2 (1-466) x US-10-208-408-18 (1-1974)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 178 TGGAAAGTAGCTGATTATCTTCCTCAGTGTGGCCCTGGGCATTGGTGCC----- 222
Qy 28 GlnLeuLeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
Db 223 -----GTTCTATAGATGATCTCTCAAGATGAGGCAACCACTGG 261
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 262 GTGCTGATCTGTCGAGGTTTCAAAATGGTGTGATATATATAGGCACCAAGGAGCGCTGC 321
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMet 87
Db 322 CATGCCCTACCAGATCATTCACCAATGGGATTCCTCGCAACAGATCGTTGTGATGATG 381
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 382 TAGCATGACATGCTTACTCTGAGACAATCCCACTCCAGGAATTTGTGATCAACAGGCC 441
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
```

```
Db 442 AATGGCACAGATGTCTATCAGGAGTCCCGAAGGACTACACTGGAGAGGATGTTACCCCA 501
Qy 128 GlnAsnPheTyrAsnValLeuGluValAsnGluSerGlyValThrGly---GlyAsnGly 146
Db 502 CAAATTTCTTGCTGTGTGAGAGGCGATCCAGACAGTGAAGGCGATAGATCCGCG 561
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
Db 562 AAGTCTGAAGAGTGGCCCGAGATCAGCTGTTCATTTACTTCACTGACCACTGATCT 621
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 622 ACTGGAATACTGGTTTTC---AATGAAGATCTTCATGTAAGGACCTGTAATGAGACC 678
Qy 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 679 ATCCATTACATGACAAACAAATGTCGGAAGATGGTGTCTACATTGAAGCTGT 738
Qy 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 739 GAGTCTGGCTCCATGATGAACAC---CTGCCGGATAACATCAATGTTATGCAACTACT 795
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
Db 796 GCTGCCAACCCAGAGAGTCTCTACGCTGTTACTATGATGAG-----840
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 841 -----AAGAGTCCACACTCTGGGGAGCTGTACAGCTCACTGGATGGAGACTCG 894
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 895 GACGTGGAAGATCTGACTAAAGAGACCTGCAAGAGCAGTACCACCTGGTAAATCGCAC 954
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 955 ACC-----AACACCGCCAGCTGTCAGTACGATGGAACAAACAACTCC 999
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 1000 ACCATGAAATGATGCAGTTTCAGGGT-----ATGAAA 1032
Qy 327 GluSerPheSerSer-----ProIleSerAsnSerGlyLeuValAsnPro 341
Db 1033 GCAGAAAGCCAGTTCTCCCGTCCCTACCTCCAGTCACACACTTGACCTACCCCCAGC 1092
Qy 342 ArgAspIleProLeuTyrLeuGluArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 1093 CCTGATGCTCTCACCATCATGAAAGGAACCTGATGACACC-----AATGATCTG 1146
Qy 362 ---GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1147 GAGGAGTCCAGGCACTCAGGAGGATCCAGCGGCATCTGGATGCGAGGCACCTCAT 1206
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1207 GAGAACTAGTCGCTGAGATGCTCTCTGTCGTCGAGCGCTCCGAGGCTGAGGTGGAGCAG 1266
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1267 CTCCTGTCGAGAGAGCC-----CCGCTCACGGGGCAC---AGCTCTACCCAGAG 1314
Qy 421 LeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrGly 436
Db 1315 GCCCTGCTGCACTTCGGGACCCACTGCTTCACTGGCAGCTCCCGCAGCTAGGATGCG 1374
Qy 437 LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr 456
Db 1375 TTGAGACATTTCTAGCTGCTGCTCAACTTTGTGAGAAGCCGTTATCGGCTTCACAGATA 1434
Qy 457 ValSerAlaIleGluGlnAlaCys 464
Db 457 valSerAlaIleGluGlnAlaCys 464
```

```
Db 1435 AATTTGTCCATGGACCACTGGTGC 1458
RESULT 6
US-10-002-600-10
; Sequence 10 Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 10
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Template ID: 1400599.2
US-10-002-600-10
Alignment Scores: 4.67e-86 Length: 2002
Pred. No.: 818.00 Matches: 179
Score: 55.22% Conservative: 80
Best Local Similarity: 38.17% Mismatches: 157
Query Match: 33.33% Indels: 53
DB: 12 Gaps: 13
US-09-934-066-2 (1-466) x US-10-002-600-10 (1-2002)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 187 TGGAAAGTAGCTCTATCTCTCAGTGTGGCCCTGGCGATTGGTGCC-----231
Qy 28 GlnLeuLeuAsnAspAsnValGluSerSerAspLysSerAlaLysGlyThrArgTrp 47
Db 232 -----GTTCTCTATAGATGATCCTGAAGATGGAGGCAACACTGG 270
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 271 GTGTGATGCTGGCAGGTTCAATGCTGTATATATATAGCCAGCAGCAGCCGTGC 330
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMet 87
Db 331 CATGCCCTACCAGATCATTCACCGCAATGGGATTCCTGACGACAGATCGTTGTGATGATG 390
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 391 TAGCATGACATTCGTTACTCTCAAGACAATCCCACTCCAGGAATTTGTGATCAACAGGCC 450
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db 451 AATGGCACAGATCTCTATCAGGAGTCCCGAAGGACTACACTGGAGAGGATGTTACCCCA 510
Qy 128 GluAsnPheTyrAsnValLeuGluGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
Db 511 CAAATTTCTCTCTGTTGAGCGGATGAGAGGAGTGAAGGCGATAGGATCCGCG 570
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAla 166
Db 571 AAGTCTGAAGAGTGGCCCGAGATCAGCTGTTCATTTACTTCACTGACCACTGATCT 630
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 631 ACTGGAATACTGTTTTC-----AATGAAGATCTTCATGTAAGGACCTCAATGACAGCC 687
```

```
Qy 187 LeuGluLysMethHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 688 ATCCATTACATGTACAAACACAAATGTTACCGAAGATGCTTCTACATTGAAGCCTGT 747
Qy 207 GluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 748 GAGCTGGGGTCCATGATGAACACAC--CTGCCGGATAAACATCAATGTTTATGCAACTACT 804
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProPro 246
Db 805 CTGCGCAACCCAGAGAGTCTCTAGCCTGTACTATGATGAG-----849
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 850 -----AAGAGGTCACCGTACCTGGGGGACTGTTACAGCGTCAACTGGATGGAAGACTCG 903
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 904 GAGCTGGAAGATCTGACTAAGAGAGCCCTGCACAGCAGTACCACCTGGTAAATCGCAC 963
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 964 ACC-----AACACAGCCAGCTGATGAGTATGGAACAAACAATCTCC 1008
Qy 307 LysAspTyrLeuSerSertyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 1009 ACCATGAAGTGTGAGTTTCAGGT-----ATGAAA 1041
Qy 327 GluSerPheSerSerProIleSerAsnSerGlyLeuValAsnPro-----341
Db 1042 CGCAAGCCAGTGTCTCCGTCCTCCCTACCTCCAGTCACACACCTGACCTCACCCCCAGC 1101
Qy 342 ArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 1102 CTTGATGTGCTCTCACCATCATGAAAGAAAGAACTGATGACACC-----AATGATCTG 1155
Qy 362 ---GluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1156 GAGGAGTCCAGGAGCTCAGGAGGAGATCCAGCGGCATCTGGATGCCAGGCACCTCATT 1215
Qy 381 AspGln-SerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLe 400
Db 1216 GAGGAAGTCACTGCGTAAGATCTCTCTGCTGCGCAGCTCCGAGGCTGAGGTGGAGCA 1275
Qy 400 uLeuThrSerThrArgThrThrGlyGlnProLeuValAspTyrAspCysPheLysTh 420
Db 1276 GCTCTCTCGAGAGAGCC-----CCGCTCAGGGGACAC---AGCTGCTACCCAGA 1323
Qy 420 rLeuValAsnSerPheLysAsnHisCys-----GlyAlaThrValHisTyrG1 436
Db 1324 GGCCCTGCTGCACTCCGGACCCACTGCTTCAACTGGCACTCCCGCCAGCTACGAGTATGC 1383
Qy 436 yLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnTh 456
Db 1384 GTTGAGACATTTGACGTGCTGTCACCTTTGTGAGAACCCGTATCCCGTTCCACAGGAT 1443
Qy 456 rValSerAlaIleGluGlnAlaCys 464
Db 1444 AANAATTCCATGGACCACGCTGTGC 1468
```

RESULT 7

```
US-09-967-796-1
; Sequence 1, Application US/09967796
; Patent No. US2002015535A1
; GENERAL INFORMATION:
; APPLICANT: Diep, Dinh
; Braxton, Scott M.
; Delegeane, Angelo M.
; TITLE OF INVENTION: A NOVEL HUMAN CYSTEINE PROTEASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
```

```
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/967,796
FILING DATE: 28-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/449,422
FILING DATE: <Unknown>
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-0048 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
```

```
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1855 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
```

```
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: 100877
LIBRARY: Adrenal
```

```
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-967-796-1
```

```
Alignment Scores:
Pred. No.: 2,17e-77 Length: 1855
Score: 744.00 Matches: 166
Percent Similarity: 54.63% Conservative: 70
Best Local Similarity: 38.43% Mismatches: 128
Query Match: 30.32% Indels: 68
DB: Gaps: 14
```

```
US-09-934-066-2 (1-466) x US-09-967-796-1 (1-1855)
Qy 8 PheGlnIleLeuValPheLeuHisAlaLeuLeuIlePheSerAlaGluSerArgLysThr 27
Db 60 TGGAAAGTAGTTGTTATCTCTCAGTGTGGCCCTGGGAATTGGTGCC-----104
Qy 28 GlnLeuLeuAsnAspAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTyr 47
Db 105 -----GTTCCTATAGATGATCTCTGAAGATGGAGCAGCAGCTGG 143
Qy 48 AlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCys 67
Db 144 GTGGTATCGTGGCAGGTTCAATGCTGCTATATATAGGCACCCAGGAGCAGCGCTGC 203
Qy 68 HisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMet 87
Db 204 CATGCCCTACCATGTTTCATTCACCCCAATGGGATTCCTGCCGACAGATCGTTGTGATTATG 263
Qy 88 TyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysPro 107
Db 264 TACGATGACATAGCTTACTCTGAAGACATCCCACTCCAGGAATTGTGATCAACAGGCC 323
Qy 108 AspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnVal 127
Db 324 AATGGCACACATGCTATCAGGAGTCCCGGAGGACTACACTGGAGAGGATGTTTACCCCA 383
Qy 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
Db 128 GlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGly 146
```

```

Db 384 CAAAATTTCCTGTTGTTGAGAGCGATCGAGAGCGATGAGGGTATAGGATCCCGC 443
Qy 147 LysValValLysSerGlyProAsnAspAsnIlePheIleTyrAlaAspHisGlyAla 166
Db 444 AAAGTCCTGAAGAGTGGTCCCGAGATCAGCTGTTCTATTATTCACCTGACCATGGATCT 503
Qy 167 ProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluVal 186
Db 504 TCTGGAATACTGTTTCCCC---AATGAAGATCTTCATGAAGGACCTGATTAGAAC 560
Qy 187 LeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCys 206
Db 561 ACCCATATACATTTCAAAACAAATGATCGAAGAGTGTCTACATGAGGCTGT 620
Qy 207 GluSerGlySerMetPheGluGlyLysLysAsnLeuAsnIleTyrAlaValThr 226
Db 621 GAGTCTGGTCCATGATGAACAC---CTGCCGATGAACATCAATGTTATGCACTACT 677
Qy 227 AlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPro 246
Db 678 GCTGCCAACCCAGAGATCTCTACGCTGTACTATGATGAG--- 722
Qy 247 ProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSer 266
Db 723 ---AAGAGTCCACGTACCTGGGGACTGTACAGCGTCAACTGGATGGAAGACTCG 776
Qy 267 AspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysArgArg 286
Db 777 GAGTGGGAAGATCTGACTAAGAGACCTGCACAAGCAGTACCACTGTTAAATCGCAC 836
Qy 287 ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeu 306
Db 837 ACC---AACACCCAGCCAGTCACTGATGGAACAAACAAATCTCC 881
Qy 307 LysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThr 326
Db 882 ACCATGAAGTGTGCGTTACGGT---ATGAAA 914
Qy 327 GluSerPheSerSer---ProIleSerAsnSerGlyLeuValAsnPro 341
Db 915 CGCAAGCCAGTCTCCCGTCCCTACCTCCAGTCAACACACTTGACCTCACCCCGCAGC 974
Qy 342 ArgAspIleProLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu 361
Db 975 CCGTATGCTCTCACCATCATGAAGAAAGAACTGATGAACACC---AATGATCTG 1028
Qy 362 GluSerLys---GluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
Db 1029 GAGGAGTCCAGGAGCTCAGGAGAGATCCAGCGGTATCTGGATGCCAGCAC--- 1082
Qy 378 LysGlnIleAspGlnSerIleThrAspIleThrAspLeuArgLeuSerValLysGlnThrAsnVal 397
Db 1083 ---CTCATCCGAGGTGAGGTGAGGACG--- 1106
Qy 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCys 417
Db 1107 ---CTCCTGTCGAGAGACCC---CGCTCAGCGGGCAC---ACCTGC 1145
Qy 418 PheLysThrLeuValAsnSerPheLysAsnHisCys 429
Db 1146 TACCCAGAGGTCTCTGTTACTTCCGACCCACTGC 1181

```

RESULT 8

```

US-10-043-487-97
: Sequence 97, Application US/10043487
: Publication No. US20030055220A1
: GENE INFORMATION:
: APPLICANT: HIRAGENICS
: APPLICANT: HIRAGENICS
: TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
: TITLE OF INVENTION: mammalian polypeptides
: FILE REFERENCE: RA778A
: CURRENT APPLICATION NUMBER: US/10/043.487

```

```

: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/261,130
: PRIOR FILING DATE: 2001-01-12
: NUMBER OF SEQ ID NOS: 561
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 97
: LENGTH: 960
: TYPE: DNA
: ORGANISM: Shigella Flexneri
: US-10-043-487-97

```

Alignment Scores:

```

Pred. No.: 1,46e-47 Length: 960
Score: 487.00 Matches: 115
Percent Similarity: 52.19% Conservative: 64
Best Local Similarity: 33.53% Mismatches: 124
Query Match: 19.85% Indels: 40
DB: 9 Gaps: 12

```

US-09-934-066-2 (1-466) x US-10-043-487-97 (1-960)

```

Qy 133 ValLeuLeuGlyAsnGluSerGlyValThrGly---GlyAsnGlyLysValValLysSer 151
Db 4 GTGTTGAGAGGGGATCGCAAGACGACTGAAGGCGCATAGGATCCGCAAGTCTCGAAGACT 63
Qy 152 GlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAla 171
Db 64 GGGCCCCAGGATCAGCTGTCTTACTTCTACTCACCATGATCTACTGCAATACTGGTT 123
Qy 172 MetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHis 191
Db 124 TTTTCCC---AATGAAGATCTTCAATGAAGCAGCTGAATGAGACCATCCATTACATGTAC 180
Qy 192 LysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMet 211
Db 181 AAACACAATGTACCGAAAGATGGTGTCTACATTGAAGCCTGTGAGTCTGGGTCCATG 240
Qy 212 PheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLys 231
Db 241 ATGAACACCAC---CTGCCGGATAACATCAATGTTTATGCAACTACTGCTGCCAACCCAGA 297
Qy 232 GluSerSerTrpGlyValTyrCysProGluSerTyrProProProSerGluIleGly 251
Db 298 GAGTCGCTCTACGCTCTTACTATGATGAG---AAGAGGTCC 336
Qy 252 ThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMet 271
Db 337 ACCTACCTGGGGGACTGGTACACGCTCACTGGATGGAAAGACTCGGACGTGAAGATCTG 396
Qy 272 SerLysGluThrLeuGluGlnTyrHisValValLysArgValClySerAspVal 291
Db 397 ACTAAAGAGACCTGCAACAGCAGTACCACCTGGTAAATCCACACC--- 444
Qy 292 ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer 311
Db 445 ---AACACACCGCCGCTCATGTCAGTATGCAAAACAAACAACTCCACCCTGAAGTGTG 501
Qy 312 SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSer 331
Db 502 CAGTTTCAGGGT---ATGAAACGCAAGCCAGTCTCT 534
Qy 332 ---ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeu 346
Db 535 CCCGTCCTCCCTACCTCCAGTCAACACCTTGACCTCACCCCGCCCTGATGTCCTCTC 594
Qy 347 LeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeu---GluSerLysGlu 365
Db 595 ACCATCATGAAAGGAACTGTATGAACACC---AATGATCTGGAGGAGTCCAGGCAG 648
Qy 366 AlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThr 385
Db 649 CTCACGGAGGATCCAGCGGATCTGATGCCAGGCACCTCATTTGAGAAGTCACTGCGT 708

```

```

Qy 386 AspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArg 405
Db 709 AAGATCGTCCTCTGCTGGCAGCGCTCCGAGGCTGAGGTGGAGCAGCTCCTGTCGAGAGA 768
Qy 406 ThrThrGlnProLeuValAspTrpAspCysPheLysThrLeuValAsnSerPhe 425
Db 769 GCC-----CGCTCACGGGCAC---AGCTGCTACCCAGAGGCCCTGCTGCACATTC 816
Qy 426 LysAsnHisCys-----GlyAlaThrValHisTrpGlyLeuLysTyrThrGly 441
Db 817 CGGACCCACTGCTTCAACTGGCACTCCCCACGTCAGGATATGCGTTCGAGACATTTGTAC 876
Qy 442 AlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGlu 461
Db 877 GTGCTGTGTCAACCTTTGTGAGAAGCCGTATCCACTTCACAGGATAAAATTTGCCATGGAC 936
Qy 462 GlnAlaCys 464
Db 937 CACGTGTGC 945

RESULT 9
US-09-878-574-4049
; Sequence 4049, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 4049
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(405)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-002-Q1-B1-C9
US-09-878-574-4049

Alignment Scores:
Pred. No.: 8.48e-44 Length: 405
Score: 450.00 Matches: 79
Percent Similarity: 74.63% Conservative: 21
Best Local Similarity: 58.96% Mismatches: 34
Query Match: 18.34% Indels: 0
DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x US-09-878-574-4049 (1-405)
Qy 123 GluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThr 142
Db 2 GAAGATGTTACTGTGTGGTAACTTTTTTGTCTGCTTTTGGAAACAAGTCGGCACTTACT 61
Qy 143 GlyGlyAsnGlyTyrValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAla 162
Db 62 GGTGGCAGTGGGAAGGTGTGGCAGGTGGACCTGATGATATATATTTGTATATATACT 121
Qy 163 AsphGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAsp 182
Db 122 GACCATGGAGTCCAGGGGTCTCGGGATGCTGCTGGTCTTACTTTATATCGGATGAT 181
Qy 183 PheAsnGluValLeuGluLysMetHisLysArgLysTyrAsnLysMetValIleTyr 202
Db 182 CTGATTGAAGTCTTGAAGAAAAACATGCTTCTGGAACGTATAAAAAACCTAGTATTTTAT 241

```


US-09-934-066-2 (1-466) x US-09-960-352-11028 (1-422)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrrpAla 48
Db ::::|||||
94 GTCTGGGAACCTGGTGTGCTCCCTGGAGGACCTCGAGGACGAGGCAAGCACTGGGTG 153
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|||||
154 GTGATCGTCGAGGATCAACAGCGGTGTATATATACAGGACCAAGGAGATGATGATGAT 213
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|||||
214 GCCTACAGATCGTTCACCAAGGATCCCTGATGAGGATCATCGTGTGATGATGAT 273
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|||||
274 GAGCAGATCGCAACTCCGAAGACAAATCCACCCCGGGAATGTGATCAACAGGCCCAAC 333
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGln 128
Db ::::|||||
334 GGCTCAGACGTGTACCAAGCGGTGTGTAAGGACTTACACGGCGGAAGATGTCAACCCCAAG 393
Qy 129 AsnPheTyrAsnValLeuLeuGlyAsn 137
Db ::::|||||
394 AATTTCCTGCTGTGTGAGAGGTGAT 420
```

RESULT 14

US-09-960-352-1596
; Sequence 1596, Application US/09960352
; Patent No. US20020137139A1

GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1596
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 07-LIB34-074-Q1-EI-B3
US-09-960-352-1596

Alignment Scores:

Pred. No.:	5.62e-30	Length:	426
Score:	333.00	Matches:	62
Percent Similarity:	69.72%	Conservative:	14
Best Local Similarity:	56.88%	Mismatches:	33
Query Match:	13.57%	Indels:	0
DB:	10	Gaps:	0

US-09-934-066-2 (1-466) x US-09-960-352-1596 (1-426)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrrpAla 48
Db ::::|||||
98 GTCTGGGAACCTGGTGTGCTCCCTGGAGGACCTCGAGGACGAGGCAAGCACTGGGTG 157
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|||||
158 GTGATCGTCGAGGATCAACAGCGGTGTATATATACAGGACCAAGGAGATGATGATGAT 217
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|||||
218 GCCTACAGATCGTTCACCAAGGATCCCTGATGAGGATCATCGTGTGATGATGAT 277
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|||||
```

```
Db 278 GAGCAGATCGCAACTCCGAAGACAAATCCACCCCGGGAATGTGATCAACAGGCCCAAC 337
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGln 128
Db ::::|||||
338 GGCTCAGACGTGTACCAAGCGGTGTCTCAAGGACTACACGGCGGAGGATGTCAACCCCAAG 397
Qy 129 AsnPheTyrAsnValLeuLeuGlyAsn 137
Db ::::|||||
398 AATTTCCTGCTGTGTGAGAGGTGAT 424
```

RESULT 15

US-09-960-352-5110
; Sequence 5110, Application US/09960352
; Patent No. US20020137139A1

GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5110
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 22-LIB34-005-Q1-EI-F5
US-09-960-352-5110

Alignment Scores:

Pred. No.:	6.05e-27	Length:	407
Score:	307.00	Matches:	55
Percent Similarity:	72.04%	Conservative:	12
Best Local Similarity:	59.14%	Mismatches:	26
Query Match:	12.51%	Indels:	0
DB:	10	Gaps:	0

US-09-934-066-2 (1-466) x US-09-960-352-5110 (1-407)

```
Qy 29 LeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThrArgTrrpAla 48
Db ::::|||||
124 GTCTGGGAACCTGGTGTGCTCCCTGGAGGACCTCGAGGACGAGGCAAGCACTGGGTG 183
Qy 49 ValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHis 68
Db ::::|||||
184 GTGATCGTCGAGGATCAACAGCGGTGTATATATACAGGACCAAGGAGATGATGATGAT 243
Qy 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db ::::|||||
244 GCCTACAGATCGTTCACCAAGGATTCCTGATGAGGACGATCATCGTGTGATGATGAT 303
Qy 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db ::::|||||
304 GAGCAGATCGCAACTCCGAAGACAAATCCACCCCGGGAATGTGATCAACAGGCCCAAC 363
Qy 109 GlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db ::::|||||
364 GGCTCAGACGTGTACCAAGGCGGTGCTGAAGGACTACACG 402
```

Search completed: May 29, 2003, 20:43:48

Job time : 209 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:21:49 ; Search time 2474 Seconds
(without alignments)
4735.812 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSSPLGHPQILVFLHALLIF.....CNMGVDVKVTSIEQACSM 466

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-O=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163701_24421/app_query.fasta_1.647
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=p2n.rnnp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MAIRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0934066_@CIGN_1_1_4447_@runat_19052003_163701_24421
-NCPU=6 -ICPU=3 -NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -VGAPOP=7 -VGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*

29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2454	100.0	1401	19	US-09-513-996A-66846
2	2454	100.0	1560	35	US-09-934-066-1
3	2454	100.0	1560	66	US-60-226-804-1
4	2446	99.7	1541	19	US-09-513-996A-11859
5	2188	89.2	4320	35	US-09-934-066-3
6	2188	89.2	4320	66	US-60-226-804-3

Sequence 66846, A
Sequence 1, Appli
Sequence 1, Appli
Sequence 11859, A
Sequence 3, Appli
Sequence 3, Appli

7 2188 89.2 83253 20 US-09-534-859-443 Sequence 443, App
8 2188 89.2 83253 31 US-09-803-736-443 Sequence 443, App
9 1287 52.4 1758 42 US-10-219-999-11010 Sequence 11010, A
10 1287 52.4 1758 42 US-10-219-999-11010 Sequence 11010, A
11 1260 51.3 1914 25 US-09-654-617-3227 Sequence 89, Appl
12 1260 51.3 1914 25 US-09-654-617-3227 Sequence 89, Appl
13 1260 51.3 1914 25 US-09-654-617-3227 Sequence 89, Appl
14 1246 50.8 1485 19 US-09-513-996A-37409 Sequence 37409, A
15 1246 50.8 1485 19 US-09-513-996A-37409 Sequence 37409, A
16 1246 50.8 1485 19 US-09-513-996A-37409 Sequence 37409, A
17 1237 50.4 1842 28 US-09-708-427-10759 Sequence 10759, A
18 1237 50.4 1842 28 US-09-708-427-10759 Sequence 10759, A
19 1228.5 50.1 1946 25 US-09-654-617-385166 Sequence 385166, A
20 1228.5 50.1 1946 25 US-09-654-617-385166 Sequence 385166, A
21 1205.5 49.1 1881 42 US-10-219-999-25725 Sequence 25725, A
22 1204.5 49.1 1786 42 US-10-219-999-26896 Sequence 26896, A
23 1204.5 49.1 1794 42 US-10-219-999-29072 Sequence 29072, A
24 1204.5 49.1 1817 42 US-10-219-999-25215 Sequence 25215, A
25 1201.5 49.0 1769 42 US-10-219-999-19189 Sequence 19189, A
26 1201.5 49.0 1774 42 US-10-219-999-12877 Sequence 12877, A
27 1198.5 48.8 1993 25 US-09-654-617-261244 Sequence 261244, A
28 1198.5 48.8 1993 25 US-09-654-617-261244 Sequence 261244, A
29 1188 48.4 1945 25 US-09-654-617-3228 Sequence 3228, A
30 1188 48.4 1945 25 US-09-654-617-3228 Sequence 3228, A
31 1166.5 47.5 1529 28 US-09-705-926-1123 Sequence 1123, A
32 1165.5 47.5 1529 28 US-09-705-926-1123 Sequence 1123, A
33 1165.5 47.5 1529 28 US-09-705-926-1123 Sequence 1123, A
34 1160 47.3 1461 35 US-09-884-016-3230 Sequence 3230, A
35 1154 47.0 1725 25 US-09-534-842A-927 Sequence 927, App
36 1154 47.0 1725 25 US-09-534-842A-927 Sequence 927, App
37 1129.5 46.0 1721 42 US-09-684-016-129998 Sequence 129998, A
38 1125.5 45.9 1809 42 US-10-219-999-30582 Sequence 30582, A
39 1100.5 44.8 1787 25 US-09-654-617-269704 Sequence 269704, A
40 1100.5 44.8 1787 25 US-09-654-617-269704 Sequence 269704, A
41 1078.5 43.9 1527 42 US-10-219-999-31085 Sequence 31085, A
42 1066 43.4 1949 25 US-09-654-617-1012 Sequence 1012, A
43 1063 43.3 1651 42 US-09-684-016-1012 Sequence 1012, A
44 1063 43.3 1651 42 US-09-684-016-1012 Sequence 1012, A
45 1007.5 41.1 1488 42 US-10-219-999-13853 Sequence 13853, A

ALIGNMENTS

RESULT 1
US-09-513-996A-66846
Sequence 66846, Application US/09513996A
GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
FILE REFERENCE: 2750-709P
CURRENT APPLICATION NUMBER: US/09/513.996A
CURRENT FILING DATE: 2000-02-25
NUMBER OF SEQ ID NOS: 81028
SEQ ID NO 66846
LENGTH: 1401
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: UNSURE
LOCATION: 1..1401
OTHER INFORMATION: any n or Xaa = unknown
FEATURE:
OTHER INFORMATION: Location 1..1401 / Ceres Seq. ID 2176034
US-09-513-996A-66846

Alignment Scores:
Pred. No.: 4.07e-234 Length: 1401
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-09-934-066-2 (1-466) x US-09-513-996A-66846 (1-1401)

Qy 1 MetSerSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuPhe 20
Db 1 ATGTCTAGTCTCTGTGCTACATCTTCAATGTCCTGTTCTTCTCATGCTTGTATCTTC 60
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLys 40
Db 61 TCAGCTGAGTCCGCCAAACCCCAATTCCTCAACGATAATCATCTTCAATCTAGCGACAAG 120
Qy 41 SerAlaLysGlyThrArgTTPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 121 AGTCAAAAGCCACACATGGCTGTTTATAGTTGCTGGATCAATCAATATATATACTAC 180
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAsp 80
Db 181 AGGATCAGGCTGACATATGCCACCGGTATCAGATACCTCCGAAGAGCGGTTTAAAGAT 240
Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 241 GAAACATCATTTGTTTATGTATGATGATCGCGTTTCTCGGAGAACTCTAGCCCT 300
Qy 101 GlyValIleLeuAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 301 GGAGTTATCATTAATAAACCCAGATGGAGAGATGTTTATAAAGAGATCTCTAAGGACTAC 360
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 361 ACTAAGAAGCTGTTAATGTTTCAAACTTCTACAATGCTTACTTGAATAAGAGTGGC 420
Qy 141 ValThrGlyLysGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 421 CTCACAGAGGAAATGGCAAGTTGTGAAAAGTGGTCTAATCATATATCTTCATCTAT 480
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 481 TATGCTGACCATGAGCTCTCGCTTAAATAGCATGCCACTGCCATGAAGTTATGGCA 540
Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db 541 AAGATTTCATGAAGTCTTGGAGAGATGCATAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 601 ATCTATGTTGAAGCATGTTGAATCAGGAAGTATGTTTGAAGGAGATTTTAAAGAGAG 660
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db 661 AACATATACGAGTACTGCTGCTAATCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy 241 GluSerTyrProProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db 721 GAGTCATATCT 780
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 781 TCTTGGCTTGAAGACAGTACCTTTCATGACATGAGCAAGAGAGAGAGAGAGAGAGAG 840
Qy 281 HisValValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 841 CACCTTGTAAAG 900
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 901 GGAACAG 960
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 961 GATAACTTCACTTTCACGGAATCTCTTCTCTCAACATCTCTAATCTCTGCTTGGTCAAT 1020
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGluLysAlaProMetGlySer 360
Db 1020 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGluLysAlaProMetGlySer 360

Db 1021 CCSCGGCATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAATGGGATCA 1080
Qy LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db CTTGAAAGCAAGAGCTCAGAGAAATCTCTTGACGAAAGAAATCATAGAAACAATC 1140
Qy AspGluSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db GATCAGAGCATTAACAGACATCTCGCGCTTTCAGTTAAACAAACCAATGCTTAAATCTC 1200
Qy LeuThrSerThrArgThrGlnGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db TTAACCTTCCAGAGCAACAGACAGCCCTCTGTAGACATTGGGATTGCTTCAAGACT 1260
Qy LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db CTAGTTAATAGCTCAAGATCACTCGGCTGCAACGGTGATTAACGATTGAAGTATACA 1320
Qy GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCAGCCATT 1380
Qy GluGluAlaCysSerMet 466
Db GAACAAGCTTGTTCGATG 1398
RESULT 2
US-09-934-066-1
: Sequence 1. Application US/09934066
: GENERAL INFORMATION:
: APPLICANT: Gruis, Rudolf
: TITLE OF INVENTION: Methods of Increasing Polypeptide
: FILE REFERENCE: 35718/237251
: CURRENT APPLICATION NUMBER: US/09/934,066
: CURRENT FILING DATE: 2001-08-21
: PRIOR APPLICATION NUMBER: US 60/226,804
: PRIOR FILING DATE: 2000-08-21
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 1560
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-934-066-1
Alignment Scores:
Pred. No.: 4,7e-234 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 35 Gaps: 0
US-09-934-066-2 (1-466) x US-09-934-066-1 (1-1560)
Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 40 ATGCTAGTCCTCTGGTCACTTTCAGATCTCTGTTTCTTCATGCTTGTCTTATCTTC 99
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 100 TCAGCTGAGTCCGCAAAACCCCAATGCTCAACGATAATGATCTGAATCTACGCAAG 159
Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 160 AGTGCAAAAGGCAACGATGGCTGTTTATGCTGGATCAATGAATATTAATACTAC 219
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 220 AGGCATCAGGCTGACATATGCCACGGGTATCAGATACTCCGAAAAGCGGCTTTAAAGAT 279

Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db GAAACATCATCTGTTTATGATGATATCGCGTTTTCTCGGAGAAATCTAGSCCT 339
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db GGAGTTTATCATTAATAAACCCAGATGGAGAGATGTTTATAAAGGAGTCTCTAAGGACTAC 399
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db ACTAAGAGAGCTGTTAATGTTCAAACTCTACAACTGTGTACTTGGAAATGAAGTGGC 459
Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db GTCACAGGAGAAATGCAAAAGTTGTGAAAGTGGTCTAATGATAATATCTTCATCTAT 519
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db TATGCTGACCATGGAGCTCTGCTTAAATAGCCATGCCACTGGTGATGAAGTTATGGCA 579
Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db AAAGATTCAATGAAGCTCTGGAGAGATGCATAAGAGAAAAAATACACAAGATGGTG 639
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db ATCTATGTTGAAGCATGTGAATCAGGAAGTATCTTTGAAGGGATTTTAAAGAAAAATCTC 699
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db AACATATACCGACGTGACTGCTGCTAATCTTAAAGAGAGCAGCTGGGAGTTTACTGCTCT 759
Qy 241 GluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db GAGTCATATCTCTCTCTCTCTGAGATGGAACTGTCTCGGCCATACATTTAGCATC 819
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db TCTTGGCTTGAGGACAGTGCCTTCATGACATGAGCAAGAGAGACTTTGGAGCAACAATC 879
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db CAGCTTGTAAAGAGAGAGTAGATCTGATGTACCAGAGACTTCTCATGTATCCGCTTTC 939
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db GGAACAGAGAGATGCTTAAAGATTATCTTCTCTTACATTCGAGAGAAATCTCGAAAC 999
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db GATAACTTCACTTTCACGGAATCCTTTCTCACCATCTCTAATCTTGGCTTGGTCAAT 1059
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db CCGCGCGATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAAGCTCCAATGGGATCA 1119
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db CTTGAAAGCAAGAGAGCTCAGAGAAATGCTTGAGCAAAAGAAATCATAGCAACAATC 1179
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db GATCAGAGCATTAACAGACATCTCGCGCTTTCAGTTAAACAAACCAATGCTTAAATCTC 1239
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db TTAACCTTCCAGAGCAACAGACAGCCCTCTGTAGACGATTGGGATTGCTTCAAGACT 1299
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db CTAGTTAATAGCTTCAAGAAATCACTCGGCTGCAACGGTGATTAAGGATTGAAGTATACA 1359
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460

Db 1360 GGAGCCCTGGCCAATATCTGCAATATGAGGATGGATGTGAAGCAAACTGTTTCAGCCATT 1419
Qy 461 GluGlnAlaCysSerMet 466
Db 1420 GAACAGCTGTTCGATG 1437

RESULT 3

US-60-226-804-1
; Sequence 1, Application US/60226804
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; TITLE OF INVENTION: Methods of Increasing Polypeptide
; FILE OF INVENTION: Accumulation in Plants
; FILE REFERENCE: 5718-134P
; CURRENT APPLICATION NUMBER: US/60/226,804
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-60-226-804-1

Alignment Scores:
Pred. No.: 4,7e-234 Length: 1560
Score: 2454.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-934-066-2 (1-466) x US-60-226-804-1 (1-1560)

Qy 1 MetSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 40 ATCTCTAGTCCCTTGGTCACTTCAGATTCTTGTTCCTTCATGCTTTCATCTTC 99
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLys 40
Db 100 TCAGCTGAGTCCGCAAAACCCCAATGCTGAACGATAATGATGTAATCTAGCGACAAG 159
Qy 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 160 AGTGCAAAAGGCACACGATGCGCTGTTTGTAGTGTGCTGATCAATGAATATATAACTAC 219
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAsp 80
Db 220 AGGCATCAGGCTGACATATGCCACGCTATCAGATACTCCCAAAAGCGGTTTAAAGAT 279
Qy 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 280 GAAACATCATTTGTATTATGATCATCATATCGCGTTTTCTCGGAGATCCTTAGCGCT 339
Qy 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 340 GGAGTTATCATTAATACACAGATCGAGAGATGTTTATAAGGAGTTCCTAAGGACTAC 399
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 400 ACTAAAGAGCTGTTAATGTTCAAACTCTACATGTTTACTGTGAATGAAGTGGC 459
Qy 141 ValThrGlyLysGlyValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 460 GTCACAGGAGGAATGGCAAGTGTGAAAGTGGTCTCAATGATAATATCTTCATCAT 519
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 520 TATGCTGACCATGGAGCTCTGCTGCTTAATAGCATGCGCCACTGCTGATGAATATGCA 579
Qy 181 LysAspPheAsnGluValLeuGluLysMethIleLysArgLysLysTyrAsnLysMetVal 200

Db 580 AAAGATTTCAATGAAGTCTTTCGAGAAGATCATTAAGAGAAAAAATACACAAGATGGTC 639
Qy 201 IleTyrValGluAlaCysGluSerClySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 640 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTCAGAGGATTTTAAAGAAAAATCTC 699
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPro 240
Db 700 AACATATACGAGTACGCTGCTTAATCTAAAGAGAGCAGCTGGGGAGTTTACTCTCCT 759
Qy 241 GluSerTyrProProProProSerGluIleClyThrCysLeuGlyAspThrPheSerIle 260
Db 760 GAGTCATATCT 819
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 820 TCTTGGCTTGAGGACAGTACCTTTCATGACATGACAAAGAGACTTTGGAGCAACAATAC 879
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 880 CAGCTTGTAAAGAGAGAGTAGGATCTGATGTACCAGACTTCTCATGTATCCCTTTC 939
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 940 GGAACAGAGAGAGATCTTAAAGATTATCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 999
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsn 340
Db 1000 GATAACTTTCATCTTTCACGGAATCTCTTCTCTCAACCAATCTCTAATCTGGCTTGGTCAAT 1059
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1060 CCGCGCATATCT 1119
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1120 CTTGAAACCAAGAGAGCTCAGAAAGAAATGCTTTCAGCAAGAAATCATAGCAAAATC 1179
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1180 GATCAGAGCATTTACAGACATTTCTCGCGCTTTCAGTTAAACAAACCAATGCTTTAAATCTC 1239
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThr 420
Db 1240 TTAACCTTCCACAAGAACCAACAGAGAGCTCTTGTAGACGATTGGGATTGCTTCAAGACT 1299
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1300 CTAGTTAATAGCTTCAAGAAATCACTGCGGTGCAACGCTGCATTACGATTTGAAGTATACA 1359
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1360 CGAGCGCTTGCCATATCTGCAATATGGAGTGTGAGCAAACTGTTTCAGCCATT 1419
Qy 461 GluGlnAlaCysSerMet 466
Db 1420 GAACAGCTTGTTCGATG 1437

RESULT 4

US-09-513-996A-11859
; Sequence 11859, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 11859
; LENGTH: 1541
; TYPE: DNA

```

: ORGANISM: Arabidopsis thaliana
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: 1..1541
: OTHER INFORMATION: any n or xaa = unknown
: FEATURE:
: OTHER INFORMATION: Location 1..1541 / Ceres Seq. ID 1375309
US-09-513-996A-11859

Alignment Scores:
Pred. No.:      2,91e-233      Length:      1541
Score:          2446.00        Matches:      465
Percent Similarity: 99.79%      Conservative: 0
Best Local Similarity: 99.79%      Mismatches: 1
Query Match:      99.67%      Indels:      0
DB:              19          Gaps:      0

US-09-934-066-2 (1-466) x US-09-513-996A-11859 (1-1541)

Qy 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
Db 59 ATGCTAGTCTCTGCTGCTACTTTCAGATTCTGTTTTCTTCATGCTTTGCTTATCTTC 118
Qy 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAsPlys 40
Db 119 TCAGCTGAGTCCGCAAAACCCAAATGCTGAACGATAATGATGTTGAATCTAGCGACAAG 178
Qy 41 SerAlaLysGlyThrArgTTPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 179 AGTCAAAAGGCACACAGCTGGCTGTTTGTAGTGTGATCAAAATGAATATTAAGTAC 238
Qy 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 239 AGGCATCAGCTGACATATGCCACGCGTATCAGATATCTCCGAAAGCGGTTTAAAGAT 298
Qy 81 GluAsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 299 GAAACATCATGTTGTTATGTATGATGATGATCGCTTTTCCCTCGGAGAAATCTTAGGCCT 358
Qy 101 GlyValIleLeuAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 359 GGAGTTATCATTAATAAACAGATGGAGAGATGTTTATAAGAGAGTTCCTAAGGACTAC 418
Qy 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 419 ACTAAGAAGCTGTTAATGTTTCAAACTTCTACAATGTGTTACTTGGAAATGAAAGTGC 478
Qy 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 479 GTACAGGAGAAATGGCAAGTTGTGAAAGTGGTCTTAATGATAATATCTTCATCTAT 538
Qy 161 TyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAla 180
Db 539 TATGCTGACCATGGAGCTCTGCTGCTTAATAGCGATGCCACTGCTGATGAAGTATGGCA 598
Qy 181 LysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetVal 200
Db 599 AAAGATTTCATGAAGTCTTGGAGAGATGATGATAAGAGAAAAATATACAAAGATGGTG 658
Qy 201 IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeu 220
Db 659 ATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGATTTTAAAGAAAAATCTC 718
Qy 221 AsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyrCysPro 240
Db 719 AACATATACCGATGACTGCTGCTAATCTTAAGAGAGAGCTGGGAGTTACTGTCTCT 778
Qy 241 GluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIle 260
Db 779 GAGTCATATCTCTCTCTCTCTCTGAGATTGGAACCTTGTCTCGCGCATACATTTAGCATC 838
Qy 261 SerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyr 280
Db 838
```

```

Db 839 TCTTGCTTGGAGACAGTGCACCTTCATGACATGAGCAAGAGACTTTTGGAGCAACAATAC 898
Qy 281 HisValValLysArgValGlySerAspValProGluThrSerHisValCysArgPhe 300
Db 899 CACGTTGTAAGAGAGAGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 958
Qy 301 GlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsn 320
Db 959 GGAACAGAGAGATGCTTAAAGATTATCTTCTCTCTTACATTTGAGAGAAATCTCTGAAAC 1018
Qy 321 AspAsnPheThrPheThrGluSerPheSerSerPheSerSerPheSerSerGlyLeuValAsn 340
Db 1019 GATAAATCTCACTTTCACGGAATCTTTTCTCTCAACATCTCTAATCTGCTTGGTCAAT 1078
Qy 341 ProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySer 360
Db 1079 CCGCGCGATATTCTCTGCTATACCTCCAGAGAAAGATTCCAAAAGCTCCAAATGGGATCA 1138
Qy 361 LeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIle 380
Db 1139 CTTGAAAGCAAGAAAGCTCAGAAAGATGCTTGACGAAAGAAATCATAGCAACAATC 1198
Qy 381 AspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeu 400
Db 1199 GATCAGAGCATTCACAGACATTCGCGCTTTCAGTTAAACAAACCAATGTCTTAAATCTC 1258
Qy 401 LeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCysPheLysThr 420
Db 1259 TTAACITCCACAGAGAACACAGACAGCCCTCTTGTAGACGATTTGGGATTTGCTTCAAGACT 1318
Qy 421 LeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThr 440
Db 1319 CTAGTTAATAGCTTCAAGAAATCACTGCGGTGCAACGGTGCATTACGGATTGAAGTATACA 1378
Qy 441 GlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIle 460
Db 1379 GGAGCGCTTCCCAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTTCAGCCATT 1438
Qy 461 GluGlnAlaCysSerMet 466
Db 1439 GAACAAGCTTGTCTGATG 1456

RESULT 5
US-09-934-066-3
: Sequence 3, Application US/09934066
: GENERAL INFORMATION:
: APPLICANT: Gruis, Darren B.
: TITLE OF INVENTION: Methods of Increasing Polypeptide
: FILE REFERENCE: 35718/237251
: CURRENT APPLICATION NUMBER: US/09/934,066
: PRIOR FILING DATE: 2001-08-21
: PRIOR APPLICATION NUMBER: US 60/226,804
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 4320
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-934-066-3

Alignment Scores:
Pred. No.:      6.43e-207      Length:      4320
Score:          2188.00        Matches:      463
Percent Similarity: 67.89%      Conservative: 0
Best Local Similarity: 67.89%      Mismatches: 3
Query Match:      89.16%      Indels:      219
DB:              35          Gaps:      7

US-09-934-066-2 (1-466) x US-09-934-066-3 (1-4320)
```


: TITLE OF INVENTION: Methods of Increasing Polypeptide

: FILE REFERENCE: 5718-134P

: CURRENT APPLICATION NUMBER: US/60/226,804

: CURRENT FILING DATE: 2000-08-21

: NUMBER OF SEQ ID NOS: 3

: SOFTWARE: FastSeq for Windows Version 3.0

: SEQ ID NO 3

: LENGTH: 4320

: TYPE: DNA

: ORGANISM: Arabidopsis thaliana

: US-60-226-804-3

Alignment Scores:

Pred. No.: 6,43e-207 Length: 4320
Score: 2188.00 Matches: 463
Percent Similarity: 67.89% Conservativeness: 0
Best Local Similarity: 67.89% Mismatches: 3
Query Match: 89.16% Indels: 219
DB: 66 Gaps: 7

US-09-934-066-2 (1-466) x US-60-226-804-3 (1-4320)

QY	1	MetSerSerProLeuGlyHisPheGlnLeuValPheLeuHisAlaLeuLeuIlePhe	20
DB	2020	ATCTAGTCTCTTGGTCACTTTTCAGATTCTTTCTTCATGCTTTGCTTATCTTC	2079
QY	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys	40
DB	2080	TCAGCTGAGTCCGCAAAACCCCAATGCTGAAGCATTAATGTTGAATCTACGACAAG	2139
QY	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTrpAsnTyr	60
DB	2140	AGTCAAAAGGCACACGATGGCTGTTTGTAGTGGATCAATGAATATTATACTAC	2199
QY	61	ArgHis-----	62
DB	2200	AGGCATCAGCTTGTAAATATTGTTGAAGCTTTAAACATAACAAAAAAGGTCCA	2259
QY	63	-----GlnAlaAspIleCysHis	68
DB	2260	AGCGAGATTGTATGAACATAAATCGACCGAGCTTTTATTTTCACAGGCTGACATGCCAC	2319
QY	69	AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleValPheMetTyr	88
DB	2320	GCATATCAGATACTCCGAAAGCGGTTTAAAGATGAAACATCATTTGTTATGAT	2379
QY	89	AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleLeuAsnLysProAsp	108
DB	2380	GATCATATCCGCTTTTCTCGGAGAAATCCCTAGGCCTGGAGTTTATCATTAATAAACCCAGAT	2439
QY	109	GlyGluAspValTyrLysGlyValProLys-----	118
DB	2440	GGAGAAGATGTTTATAAGAGATGTTCTTAA-----GGTCTTATTTCATCTTTTGTGCGTTA	2498
QY	118	-----	118
DB	2499	TTTCTAGCTTGAATTCATATATATATATATATTCACAGTTTGTGTTTATTGTTGGGTAG	2558
QY	119	AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu	138
DB	2559	GACTACACTAAGAAGCTGTTAATGTTTCAAAACTTCTACAAATGTTTACTTGGAAATGAA	2618
QY	139	SerGlyValThrGlyLysAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe	158
DB	2619	AGTGGCTCACAGGAGAAATGCCAAGTTGTGAAGAAGTGGTCTTAATGATATATCTTC	2678
QY	159	IleTyrTrpAlaAspHisGlyAlaProGlyLeuIle-----	170
DB	2679	ATCTATTATGCTGACCATGGAGCTCCTGGCTTAATAGGTTTCTTAATTTTATGAAATTA	2738
QY	170	-----	170

DB	2739	TTACGTACCATCAATCCATATATATATAAAGATTCTTCTTTGATACACTACGAAACCGCG	2798
QY	171	-----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu	187
DB	2799	ATTTTCTCAGCGATGCCACTGCTGATGAAGTTATGGCAAAAGATTTCATGAAGTCTTG	2858
QY	188	GluLysMetHisLysArgLysLysTyrAsnLysMetVal-----	200
DB	2859	GAGAAGATGCATAAGAGAAAAAATACAAACAGATGGT-ATATAACTCAACCATTCGTGA	2917
QY	200	-----	200
DB	2918	CCTAGCTTTATACATATGCTTCTGTTTTCATCTCTATGCTGCTGTTTTCGATGTT	2977
QY	201	-----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys	218
DB	2978	TAGCTGATCTATGTTGAGCATGTGAATCAGGAAGTATGTTTGAAGGATTTTAAAGAAA	3037
QY	219	AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr	238
DB	3038	AATCTCAACATATACGACGAGTCTGCTGAATTTCTAAAGAGACGAGCTGGGAGTTTAC	3097
QY	239	CysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPhe	258
DB	3098	TGCTCCTGAGTCATATCTCTCTCTCTCTCTGAGATTGGAACCTTCTCTCGCGATACATTT	3157
QY	259	SerIleSerTrpLeuGlu-----	264
DB	3158	AGCATCTCTGGCTTGAGGACAGGTACTGCAACAAAAAAGATTCAATCCTTATGGACTA	3217
QY	264	-----	264
DB	3218	TTCCAATGATTGATTGTTCTTGAAGATATTTGTTTCATTTGTTCTATGTTTGTCTGT	3277
QY	265	-----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisVa	282
DB	3278	GTTTGGGACAGTCACCTTCATGACATGACAAAGAGACTTTTGAGCAACAATACCACGT	3337
QY	282	lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyTh	302
DB	3338	TGTAAAGAGAAGATAGGATCTCATGTACCAGAGACTTCTCATGTATGCCGTTTCGGAAC	3397
QY	302	rGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAs	322
DB	3398	AGAGAAGATCCTTAAAGATTATCTTCTCTTACATTGGAAGAAATCCTCAAAACGATAA	3457
QY	322	nPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProAr	342
DB	3458	CTTCACTTTTCACGGAATCTCTTCTCACCATACTCTAATTTCTGGCTTGGTCAATCCGCG	3517
QY	342	gAspIleProLeuLeuTyrLeuGlnArgLysIle-----	353
DB	3518	CGATATTCCTCTGCTATACCTCCAGAGAAAGGT-GAGCTTTTTCGGGTTTTTTTGATCAT	3576
QY	354	-----GlnLysAl	356
DB	3577	TTTAAACGAAAGAGTTTTCAGCATGTTTAAATGTTTATTCATCTCTTAGATTCAAAAGC	3636
QY	356	aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHi	376
DB	3637	TCCAAATGGGATCACTTGAAGCAAGAAAGCTCAGAAGAAATTCCTGACGAAAGAAATCA	3696
QY	376	sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs	396
DB	3697	TAGGAACAATCGATCAGACAGCATTTACAGACATTTCTGCGCTTTTCAGTTTAAACAAACAA	3756
QY	396	nValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAs	416
DB	3757	TGCTTTAAATCTCTTAATCTCCACAAGAACAACAGGACGCCCTCTTGTAGACGATTGGGA	3816
QY	416	pCysPheLysThrLeu-----	421
DB	3817	TTGCTTCAAGACTCTAGTAACAAAAACCAATCTCAAAACCTTTGTTACTTGTGTTCTACGAA	3876

```
QY 422 -----ValAsnSerPh 425
Db 3877 CAACATTGCATTATTACTAAACAGCTGTATATCGAATGAAATCGCAGCTTAATAGCTT 3936
QY 425 eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs 445
Db 3937 CAAGAATCACTCGCGTCAACGGTGCATTACGGATTGAAGTATACAGAGCGCTTGCCAA 3996
QY 445 nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGlnAlaCysSe 465
Db 3997 TATCTGCAATATGGAGTGGAGTGAAGCAACAACTGTTTCGCCCAATGAACAGCTTGTTTC 4056
QY 465 rMet 466
Db 4057 GATG 4060

RESULT 7
US-09-534-859-443
: Sequence 443, Application US/09534859
: GENERAL INFORMATION:
: APPLICANT: Bush, David F.
: APPLICANT: Last, Robert L.
: APPLICANT: Levin, Irena M.
: APPLICANT: Norris, Susan R.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Rounsley, Steven D.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: PLANT POLYMORPHIC MARKERS AND USES THEREOF
: FILE REFERENCE: 38-10(15493)B
: CURRENT APPLICATION NUMBER: US/09/534,859
: CURRENT FILING DATE: 2000-03-29
: NUMBER OF SEQ ID NOS: 1127
: SEQ ID NO 443
: LENGTH: 83253
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-534-859-443

Alignment Scores:
Pred. No.: 3,266-205 Length: 83253
Score: 2188.00 Matches: 463
Percent Similarity: 67.89% Conservatives: 0
Best Local Similarity: 67.89% Mismatches: 3
Query Match: 89.16% Indels: 219
DB: 20 Gaps: 7

US-09-934-066-2 (1-466) x US-09-534-859-443 (1-83253)
QY 1 MetSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuIlePhe 20
Db 72700 ATGCTAGTCTCTTGGTCACTTTCAGATCTCTGTTTCTTCATGCTTTGCTTATCTTC 72759
QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 72760 TCAGCTGAGTCCCGCAAAACCCAAATGCTGAACCAATGAATGTTGAATCTAGCGACAAG 72819
QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTrpAsnTyr 60
Db 72820 AGTGCAAAAGGCACACGATGGCTGTTTGTAGTCTGCTGATCAATGAATATTATACTAC 72879
QY 61 ArgHis----- 62
Db 72880 AGGCATCAGGTTGTTAAATTTATGTTGAACGTTTAAACATAACAAAAAAGGTCCA 72939
QY 63 -----GlnAlaAspIleCysHis 68
Db 72940 AGCGAGATTGTTGTAATAAATCGACCGAGCTTTTATTTTCACAGGCTGACATATGCCAC 72999
QY 69 AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr 88
Db 73000 GCGTATCAGATACCTCGAAAGCGGTTTAAAGATCAACATCATGTTGTTTATGATAT 73059
```

```
QY 89 AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp 108
Db 73060 GATGATATCGCGTTTCTCGGAGAATCCTAGGCTGGAGTTATCATTAATAAACCAT 73119
QY 109 GlyGluAspValTyrLysGlyValProLys----- 118
Db 73120 GGAGAAGATGTTTATAAAGAGTTCCCTAA--GGTCTTATTCTTCTCTTTCTGCGTTA 73178
QY 118 ----- 118
Db 73179 TTTCTACCTTCAATTCATATATATATATATATATATATATATATATATATATATATAT 73238
QY 119 AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu 138
Db 73239 GACTACACTAAAGAGCTGTTTAAATGTTTCAAAACTTCTACAAATGTTACTTGGAAATGAA 73298
QY 139 SerGlyValThrGlyLysValValLysSerGlyProAsnAspAsnIlePhe 158
Db 73299 AGTGGCGTCAACAGGAGGAAATGCGCAAAAGTTCTGAAAAGTGGTCTTAATGATATATCTTC 73358
QY 159 IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle----- 170
Db 73359 ATCTATTATGCTGACCATGGAGCTCCTGGCTTAAATAGGTTTCTTAATTTTATGAAATTA 73418
QY 170 ----- 170
Db 73419 TTACGTACCATCAATCCATATATATATAAAGATTTTCTTTGATACATACGAAACCGG 73478
QY 171 -----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu 187
Db 73479 ATTTTCTCAGCGATGCCACCTGGTGAATGAGATTTGCAAAAGATTTCAATGAAGTCTTG 73538
QY 188 GluLysMetHisLysArgLysLysTyrAsnLysMetVal----- 200
Db 73539 GAGAAGATGCTAAGAGAAATAAATAACAACAGATGGT--ATATAACTCAACCATTCCTTTA 73597
QY 200 ----- 200
Db 73598 CCTAGCTTTATACATATGCTGTTTCTGTTTGTGAATCTCTATGCTGCTGTTTGTGATGTT 73657
QY 201 -----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys 218
Db 73658 TAGGTGATCTATGTTGAAGCATGTGAATCAGGAAGTATGTTTCAAGGATTTTAAAGAAA 73717
QY 219 AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerTrpGlyValTyr 238
Db 73718 AATCTCAACATATACGCAGTGACTGCTGCTAAATTTCTAAACAGAGCAGCTGGGAGTTTAC 73777
QY 239 CysProGluSerTyrProProProSerCysIleGlyThrCysLeuGlyAspThrPhe 258
Db 73778 TGTCTGAGTCAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 73837
QY 259 SerIleSerTrpLeuGlu----- 264
Db 73838 AGCATCTCTTGGCTTGAGGACAGGTACTGCAACAAAAAGATTCAATCCTTTATGSACTA 73897
QY 264 ----- 264
Db 73898 TTGCAATGATTGATTGTTTCTTGAAGAATATTTGTTTCAATTTGTTCTATGTTTGTGCTGT 73957
QY 265 -----AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisVa 282
Db 73958 GTTTGGGACAGTCACCTTTCATGACATGAGCAAAAGACATTTTGGAGCAACAAATACCAGT 74017
QY 282 lValLysArgArgValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
Db 74018 TGTAAAGAGAAGACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 74077
QY 302 rGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAspProGluAsnAspAs 322
Db 74078 AGAAGATGCTTAAAGATTTCTTTCTCTTACATTGGGAAGAAATCTCTGAAACCATTA 74137
QY 322 nPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProAr 342
```

Db	74138	CTTCACTTTCCGGAATCCTTTTCTCCACCAATCTCTAATCTCGCTTGGTCAATCCGG	74197
Qy	342	gAspIleProLeuLeuTyrLeuGlnArgLysIle	353
Db	74198	CGATATCTCTGCTATACCTCCAGAGAAGGT-GAGCTTTTTCGGGTTTTTGTGATCAT	74256
Qy	354	-----GlnLysAl 356	
Db	74257	TTTAAACGAAAGAGTTTTTCAGCATGTTTTTAATGTTTTTATTCATCTCTTAGATTCAAAAGC	74316
Qy	356	aProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHi	376
Db	74317	TCAATGGGATCACCTTGAAGCAAGCAAGCTCAGAAAGAAATGCTTGGCAAGAAATCA	74376
Qy	376	sArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAs	396
Db	74377	TAGGAACAAATCGATCAGACGATTACACACATCTCGCGCTTTCAGTTAAACAAACCA	74436
Qy	396	nValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTrpAs	416
Db	74437	TGCTTTAAATCTCTTAACCTCCACAAGAACACAGGACAGCCTCTTGTAGACGATTGGGA	74496
Qy	416	pCysPheLysThrLeu-	421
Db	74497	TTGCTTCAAGACTCTAGTAACAAACACATCTCAAAACCTTGTTACTGTGTTCTACGCCAA	74556
Qy	422	-----ValAsnSerPh 425	
Db	74557	CAACCATTCGATTATTACTAAACCACTGATATCGAATCGAAATCCAGGTTAATAGCTT	74616
Qy	425	eLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAs	445
Db	74617	CAAGAACTACTCGGGTGCACACGGTGCATTAACGATTGAAGTATACAGGAGCGCTTGCCAA	74676
Qy	445	nIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluAlaCysSe	465
Db	74677	TATCTCCATATGGGAGTGATGTGAAGCAAACTGTTTCAGCCATTGAACAGCTTGTTTC	74736
Qy	465	rMet 466	
Db	74737	GATG 74740	

RESULT 8

US-09-803-736-443
 ; Sequence 443, Application US/09803736
 ; GENERAL INFORMATION:
 ; APPLICANT: Bush, David F.
 ; APPLICANT: Levin, Irena M.
 ; APPLICANT: Norris, Susan R.
 ; APPLICANT: Rounsley, Steven D.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
 ; FILE REFERENCE: 38-10(15493)D
 ; CURRENT APPLICATION NUMBER: US/09/803,736
 ; CURRENT FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 09/534,859
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
 ; PRIOR FILING DATE: 2000-10-20
 ; NUMBER OF SEQ ID NOS: 1582
 ; SEQ ID NO 443
 ; LENGTH: 83253
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-803-736-443

Alignment Scores:			
Pred. No.:	3,26e-205	Length:	83253
Score:	2188.00	Matches:	463
Percent Similarity:	67.89%	Conservative:	0
Best Local Similarity:	67.89%	Mismatches:	3
Query Match:	89.16%	Indels:	219

DB:	31	Gaps:	7
US-09-934-066-2 (1-466) x US-09-803-736-443 (1-83253)			
Qy	1	MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20
Db	72700	ATGCTCTAGTCCCTCTTGGTCCACTTTCAGATTCCTGTTTTCTTCATGCTTTGCTTATCTTC	72759
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnAspValIgluSerSerAspLys	40
Db	72760	TCAGCTGAGTCCGCAAAACCCCAATTCGTGAAGATAATGATGTTCAATCTAGCGACAAG	72819
Qy	41	SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr	60
Db	72820	AGTGGAAAAGGCACACGATGGCTGTTTTAGTTGCTGGATCAATGAATATATAACTAC	72879
Qy	61	ArgHis-	62
Db	72880	AGCATCAGGTTGTTAAATTATGTTTGAACGTTTAAACATAACAAAAAAGGTCCA	72939
Qy	63	-----GlnAlaAspIleCysHis	68
Db	72940	ACGAGATTGTTATGAACATAAATCGACCGCGCTTTTATTTTCACAGGCTGACATATCCAC	72999
Qy	69	AlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleIleValPheMetTyr	88
Db	73000	GGGTATCAGATCTCCGAAAAGCGGTTTAAAGATGAAACATCATCTGTTTATGTAT	73059
Qy	89	AspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAsp	108
Db	73060	GATGATATCGGTTTCTCGGAGAATCTCGCCTGGAGTTATCATTAATAAACCCAGAT	73119
Qy	109	GlyGluAspValTyrLysGlyValProLys-	118
Db	73120	GGAGAAGATGTTTATAAAGAGGTTCCCTAA-GGTTCTTATTCTCTACTCTCTTCTGCGTTA	73178
Qy	118	-----	118
Db	73179	TTTCTACGTTGAATTCAATTACATATATATTTCAAGTTTGTGTTGTTGGGTAG	73238
Qy	119	AspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGlu	138
Db	73239	GACTACACTAAAGAACCTGTTAATGTTCAAACTTCTCAATGTGTTACTTGGAAATGAA	73298
Qy	139	SerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePhe	158
Db	73299	AGTGGCGTCCAGGAGGAAATGGCAAGTTGTGAAAGTGGTCTTAATGATAATATCTTC	73358
Qy	159	IleTyrTyrAlaAspHisGlyAlaProGlyLeuIle-	170
Db	73359	ATCTATTATGCTGACCATGGAGCTCCTGCTTAAATAGGTTTTTCTTAATTTTATGAAATTA	73418
Qy	170	-----	170
Db	73419	TTACGTACCATCAATCCATATCTATAATAAAGATTTTCTCTTGACTACTACGAAACCGG	73478
Qy	171	-----AlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeu	187
Db	73479	ATTTTCTCAGCATGCCACTGGTGATGAAGTTATGGCAAAAGATTTCAATGAAGCTTTC	73538
Qy	188	GluLysMetHisLysArgLysLysTyrAsnLysMetVal-	200
Db	73539	GAGAAGATGCATAAGAAAAAATAACAAAGATGTT-ATATACTCAACCATTCCTGTTA	73597
Qy	200	-----	200
Db	73598	CCTAGCTTTATACATATGTTGTTCTGTTTTTGAATCTCTATGTTGTTGTTTGTGATGTT	73657
Qy	201	-----IleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLys	218
Db	73658	TAGGTGATCTATGTTGAACCATGTGAATCAGGAAGTATCTTTCAGGGATTTTAAAGAA	73717
Qy	219	AsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyr	238


```

QY 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnThrHis 281
Db 787 TGGATGGAAGACAGTGCATACACAATTTGCAACAGAACTTTACATCAACAATACGAA 846
QY 282 ValValLysArgArgVal-----GlySerAspValProGluThrSerHisValCysArg 299
Db 847 TTGGTCAACAAGACACTAGTAATGGAATCAATT--TATGGTTCCACGTGATGCAG 903
QY 300 PheGlyThrGluLysMetLeuLysAspTyrLysSerTyrIleGlyArgAsnProGlu 319
Db 904 TATGGTGCATAGGCTTAGCGAGAACATCTCGCTTATATTGGGTACAAATCCTGCT 963
QY 320 AsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuVal 339
Db 964 AATGATAATTTTACTTTTGTGCTTAA--AACTCATTTGGTGCCACCTTCAAAAGCAGTC 1020
QY 340 AsnProAspAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGly 359
Db 1021 AACCAACGTGTCAGACTCTATCCATTTTGGGATAGTTCCGCAAGCTCTGTGGGT 1080
QY 360 SerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGln 379
Db 1081 TCTTTCAGAAAGCTGACAGCTGAGAACAAATCTGTGAAGCAATGCTCTCACAGAAATGCAT 1140
QY 380 IleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsn 399
Db 1141 ATAGATGACAGCATGAAACGTAATTGGAAGCTCTTCTTGGCATTTGAAAGGGTCCAGAA 1200
QY 400 LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLys 419
Db 1201 CTGCTTAGCAGTGTAGACCTGCTGGCAACCACTTGTGTGATGACTGGGACTGCTTAAA 1260
QY 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 1261 ACATTTGGTTAGACTTTTGAGACACATTTGTGATCCCTCTCAGTATGGGATCAAAACAT 1320
QY 440 ThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAla 459
Db 1321 ATGAGTCTCTTGGCAACTTCTGCAACGCTGGAATACGAAAGAGCAAAATGGCTGAGGCC 1380
QY 460 IleGluGlnAlaCys 464
Db 1381 TCAGCACAAAGCATGT 1395

RESULT 11
: Sequence 3227, Application US/09654617
: GENERAL INFORMATION:
: APPLICANT: Kovalic, David K.
: TITLE OF INVENTION: Annotated Plant Genes
: FILE REFERENCE: 38-21(15097)D
: CURRENT APPLICATION NUMBER: US/09/654,617
: CURRENT FILING DATE: 2000-09-05
: NUMBER OF SEQ ID NOS: 463173
: SEQ ID NO 3227
: LENGTH: 1914
: TYPE: DNA
: ORGANISM: Glycine max
US-09-654-617-3227

Alignment Scores:
Pred. No.: 9,21e-115 Length: 1914
Score: 1260.00 Matches: 247
Percent Similarity: 66.39% Conservatives: 73
Best Local Similarity: 51.24% Mismatches: 137
Query Match: 51.34% Indels: 25
DB: 25 Gaps: 6

US-09-934-066-2 (1-466) x US-09-654-617-3227 (1-1914)
QY 2 SerSerProLeuGlyHis-----PheGlnIleLeuVal 12

```

```

Db 111 TCTTCTCCAATCTCACACCCATCATGGACCGTTTCCGATCCTCTCTTCTCTCGTCGGCACCC 170
QY 13 PheLeu-----HisAlaLeuLeuLeuPheSerAlaGluSerArg 25
Db 171 TTTATCATCCCTCCCTCCGGTCCCGCCACGATATTTCTCCGGTTACCTCCGAAAGCTTCC 230
QY 26 LysThrGlnLeuLeuAsnAspValGluSerSerAspLysSerAlaLysGlyThr 45
Db 231 AGGTTCTTCAAAAGCACCTGCTAAT-----GCCGATCAAAAGCATGATGGGCACC 278
QY 46 ArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHis-GlnAlaAs 65
Db 279 AGGTGGCGGCTTTAGTTCCCGTTCCAAATGGCTACTGGAATTACAGCCACCCACTCTGA 338
QY 65 PileCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIleLeuVal 85
Db 339 TGTGTCCCATGCTATCAACTACTGAGGAAGGTGCTCAAGAAGAAATAATTGTGT 398
QY 85 PheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleLeuAs 105
Db 399 ATTTATGTATGATGACATTTCTTCAACGAGAACCCCGCAGCTGGAGTCATATTATAA 458
QY 105 nLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaVal 125
Db 459 CAGTCCACATGGAATGATGTTTACAGGGGAGTCCCTAAGGATTTACATTTGCTGAAGATGT 518
QY 125 lAsnValGlnAsnPheTyrAsnValLeuLeuGluGlyAsnGluSerGlyValThrGlyGlyAs 145
Db 519 AACTGTGGCAACTTTTGTGCTGCTATCTTGGAAATTAAGTCAGCTCTTACTGCTGCAG 578
QY 145 nGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisG 165
Db 579 TGGGAAGGTTGTGGATAGTAGTGGCCCAATGATATTTATATATATCTACTCTGATCATGG 638
QY 165 yAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnG 185
Db 639 CGGTGGCGGAGTCTAGGATGCTTCTTGGAACTTATAAAGCCTAGTATTTATCTAGAGGC 698
QY 185 uValLeuGluGlyMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAl 205
Db 699 ACTCTTGAAGAAGACATGCTTCTTGGAACTTATAAAGCCTAGTATTTATCTAGAGGC 758
QY 205 aCysGluSerGlySerMetPheGluGlyLysLysAsnLeuAsnIleTyrAlaVal 225
Db 759 ATGTGNAATCTGGAGTATCTTTGAAGTCTTCTTCCAGAGGCTCTGAATATCTATGCAAC 818
QY 225 lThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProPr 245
Db 819 AACAGCTTCAATGCTGAAGAAGACAGTTGGGAACATATTGCTGGGAGTAGTCTAG 878
QY 245 oProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAs 265
Db 879 TCTCTCCCTTGAATATGAACCTGCTCGGTGACCTGTACAGTGTGCTTGGATGGAAGA 938
QY 265 pSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysAr 285
Db 939 TAGTGACATACAAATTTCCGAACAGAACTTTACATCAACAATACGACTTTGGTCAAGA 998
QY 285 gArgVal-----GlySerAspValProGluThrSerHisValCysArgPheGlyThrG 303
Db 999 AAGGACTATGAATGGAATTTCAATC---TATGGTTTCCACCTGATGAGTATGCTGACAT 1055
QY 303 uLysMetLeuLysAspTyrLeuSerSerTrpIleArgAsnProGluAsnAspAsnPh 323
Db 1056 AGGCTTAGCAAGAACAACTCTTGTATTATTTGGGTACAAATCCTCTAATAATAATT 1115
QY 323 eThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAs 343
Db 1116 TACTTTTGTGCTTAAACACTCATTG---GTCCACCTTCAAAACAGCTCAACCACTGA 1172
QY 343 pileProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSe 363

```

```

Db      1173  TGCAGATCTCATCTTCCTGGGATAGATTCCCGAAGCTCCCTGGTGGGTTCTCTTAGGAA 1232
Qy      363  rLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSe 383
Db      1233  AGCTGCAGCTGAGAAAGAAATCTGGAAGCAATGCTCTCAGAGATGCATATAGATGACAA 1292
Qy      383  rIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSe 403
Db      1293  CATGAAACTTATTGGAAAGCTCTTATTGGCATTTGAAAGGGTCCAGAACTGCTTAGCAG 1352
Qy      403  rThrArgThrThrGlnProLeuValAspAspTTPAspCysPheLysThrLysThrValAs 423
Db      1353  TGTAGACCTGCTGGGCAACCACTGTTGATGACTGGGACTGCTTAAACATTTGGTTAG 1412
Qy      423  nSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLe 443
Db      1413  GACTTTTGAGACACATTTGTGATCCCTGCTCAGTATGGGATGAAACATATGAGGTCCT 1472
Qy      443  uAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluAla 463
Db      1473  TGCAAACTTCTGCAACGCTGGAATACGAAAGAGCAAAATGGCTGAGGCTTCAGCACAAG 1532
Qy      463  aCys 464
Db      1533  ATGT 1536

RESULT 12
US-09-684-016-3227
; Sequence 3227, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 3227
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Glycine max
US-09-684-016-3227

Alignment Scores:
Pred. No.: 9.21e-115 Length: 1914
Score: 1260.00 Matches: 247
Percent Similarity: 66.39% Conservative: 73
Best Local Similarity: 51.24% Mismatches: 137
Query Match: 51.34% Indels: 25
DB: 27 Gaps: 6

US-09-934-066-2 (1-466) x US-09-684-016-3227 (1-1914)
Qy      2  SerSerProLeuGlyHis-----PheGlnIleLeuVal 12
Db      111  TCTTCTCCAATCTCACACCCATCATGACCGGTTCCGATPCCTCTTCTCGTGGCACCC 170
Qy      13  PheLeu-----HisAlaLeuLeuIlePheSerAlaGluSerArg 25
Db      171  TTATACCCCTCGCTCCGTCGCGCCAGCATATTTCTCGGTTACCTCCGAGGCTCC 230
Qy      26  LysThrGlnLeuLeuAsnAspValIgluSerSerAspLysSerAlaLysGlyThr 45
Db      231  AGGTCTTCAACACACCTGCTAAT-----GCCGATCAACAGATGAGGCGACC 278
Qy      46  ArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHis-GlnAlaAs 65
Db      279  AGGTGGGCGGCTTTAGTTGCGGTTCCCAATGGCTACTGGAATACAGGCAACCCAGTCTGA 338
Qy      65  pIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleLeva 85

```

```

Db      339  TGTTCGCATCGCATCACTACTAGGAAAGTGGTCTCAAAAGAGAAAATATTGTTGT 398
Qy      85  lPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAs 105
Db      399  ATTATGTATGATGACATTCCTTCAACGAAGAGAACCCCGCAGCTGGAGATCATTTAA 458
Qy      105  nLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaVa 125
Db      459  CAGTCCACATGGAAGATGATGTTTACAAGGGAGTCCCTAAGGATTACATTGTTGGTGA 518
Qy      125  lAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerClyValThrGlyGlyAs 145
Db      519  AACTGTTGGCAACTTTTTCCTCTATATCTGGAATAAGTACAGCTCTTACTGGTGGCAG 578
Qy      145  nGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisG 165
Db      579  TGGGAAGTTGTGATAGTGGCCCCAATGATCATATATTTATATACTACTCTCATCATGG 638
Qy      165  yAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnG 185
Db      639  CGCTCCGGAGTGTCTAGGATGCTACTAATATCCATACATGATGATCCATCCGATGATTGA 698
Qy      185  uValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGluAl 205
Db      699  AGTCTTGAAGAAGAGCATGCTTCTGGAACCTTATAAAGCCTAGTATTTTCTAGAGGC 758
Qy      205  aCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaVa 225
Db      759  ATCTGAATCTGGGAGTATCTTTCAGAGTCTCTTCCAGAAGGCTCGAATATCTATGCAAC 818
Qy      225  lThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysProGluSerTyrProPr 245
Db      819  AACAGCTTCAATTCCTGAAGAAAGCAGTTGGGAACATATATTGCTCGGGAGTATCTCTAG 878
Qy      245  oProProSerGluIleGlyThrCysLysGlyAspThrPheSerIleSerThrLysGluAs 265
Db      879  TCCTCCCTCCCTGAATATGAAACCTGCGCTGGGTGACCTGACAGTGTGGATGGGAAGA 938
Qy      265  pSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLysAr 285
Db      939  TAGTGACATACACAATTTGGGACAGAACTTTTACATCAACATACGACATGCTGTCAAAGA 998
Qy      285  gArgVal-----GlySerAspValProGluThrSerHisValCysArgPheGlyThrG 303
Db      999  AAGGACTATGAATGGAAATTCATCT--TATGGTTCCACGCTGATGTCAGTATGTCGATCAT 1055
Qy      303  uLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPh 323
Db      1056  AGGCTTAGCAAGAACAACTCTCTCTTATATTTGGGTACAAATCTCTGCTAATGATAATTT 1115
Qy      323  eThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuValAsnProArgAs 343
Db      1116  TACTTTTGTGCTTAAACACTCATTTG--GTGCCACCTTCAAAAGCAGTCAACCAACGTGA 1172
Qy      343  pIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetClySerLeuGluSe 363
Db      1173  TGCAGATCTCATCCATTTCTGGGTAAGTTCCGCAAGGCTCCTGCTGGGTTCTTCTTAGAA 1232
Qy      363  rLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSe 383
Db      1233  AGCTGAGCTGAGAAAGAAATCTTGGAAAGCAATGTCTCACAGAAATGATATAGATGACAA 1292
Qy      383  rIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSe 403
Db      1293  CATGAAACTTATTGGAAAGCTCTTATTGGCATTGAAAGGGTCCAGAACTGCTTAGCAG 1352
Qy      403  rThrArgThrThrGlyGlnProLeuValAspAspTTPAspCysPheLysThrLysLeuVal 423
Db      1353  TGTAGACCTGCTGGGCAACCACTTGTGATGACTGGGACTGCTCTTAAACATTTGGTTAG 1412
Qy      423  nSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLe 443
Db      1413  GACTTTTGAGACACATTTGTGGATCCCTGCTCAGTATGGGATGAAACATATAGAGGCTCTT 1472

```

QY 443 ualaAsnileCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGlnAl 463
 Db 1473 TGAACACTTCTGCAACGCTGGAATACGAAAGAGCAATAGCTGAGGCTTCAGCACAGC 1532

QY 463 aCys 464

Db 1533 ATGT 1536

RESULT 13

US-09-513-996A-37409
 ; Sequence 37409, Application US/09513996A
 ; GENERAL INFORMATION:
 ; APPLICANT: N. ALEXANDROV et al.
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
 ; FILE OF INVENTION: ENCODED THEREBY
 ; FILE REFERENCE: 2750-709P
 ; CURRENT APPLICATION NUMBER: US/09/513,996A
 ; CURRENT FILING DATE: 2000-02-25
 ; NUMBER OF SEQ ID NOS: 81028
 ; SEQ ID NO 37409
 ; LENGTH: 1485
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: 1..1485
 ; OTHER INFORMATION: any n or xaa = unknown
 ; FEATURE
 ; OTHER INFORMATION: Location 1..1485 / Ceres Seq. ID 1821917

US-09-513-996A-37409

Alignment Scores:
 Pred. No.: 1,648-113 Length: 1485
 Score: 1246.00 Matches: 233
 Percent Similarity: 69.89% Conservative: 71
 Best Local Similarity: 53.58% Mismatches: 121
 Query Match: 50.77% Indels: 10
 DB: 19 Gaps: 4

US-09-934-066-2 (1-466) x US-09-513-996A-37409 (1-1485)

QY 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
 Db 145 GAAACGAGCAGGAGGATCTTAACCTCCGGTACTAGTGGGCTGTCTAGTCGCGGATCTAGC 204
 QY 56 GluTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
 Db 205 GGATATTGGAATACAGGCATCAGGCTCATATATGCCATGCTCACTCACTTCTCAGGAAA 264
 QY 76 GlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAspPheAlaPheSerSer 95
 Db 265 GTGGATTCAAAGAGAGCAATATTGTGCTATTCTATGATGATGATGATGATGATGATGATGAT 324
 QY 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGly 115
 Db 325 GAGATCCAAAGCCTGGAAACCATTAATCAACAGCCCTCATGAAAGATGCTATCAAGGA 384
 QY 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeu 135
 Db 385 GTTCCAGAGATATCTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
 QY 136 GlyAsnGluSerGlyValThrGlyAsnGlyLysValLysSerGlyProAsnAsp 155
 Db 445 GGAGACAAACCTGCTTTAAAGGGAAGTGGGAGGTTGGGATAGTGGTCTTAATGAT 504
 QY 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
 Db 505 CATATCTTCATATCTACAGTACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564
 QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluTyrMetHisLysArgLysLys 195
 Db 565 CCTTACCTATATGCAATGATCTCAATGATGCTTGAAGAGAAACATGCTTTAGGAACA 624

QY 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyLeu 215
 Db 625 TATAAAGCTTGGTGTATATCTCGAAGCTTCGAATCTCGAAGTATCTTTGAAGGCTT 684
 QY 216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerTrp 235
 Db 685 CTTCCTGAGGTTTGAACATCTATGCCCAACTGCATCAACCGGAGCAAGAGAGAGTGG 744
 QY 236 GlyValTyrCysProGluSerTyrProProProSerGlyGlyGlyGlyGlyGly 255
 Db 745 GGTACCTATTGCTCGAGAGAGCAACCTCCACCGGAGTATGAAACTTGTGTAGGT 804
 QY 256 AspThrPheSerIleSerTrpLeuGluAspSerLeuHisAspMetSerLysGluThr 275
 Db 805 GACTTGTACAGTGTGCTGGATGCAAGATAGTGTATGACAAATTTACACACTGACACT 864
 QY 276 LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro 292
 Db 865 CTGCACCAATATCAACTTCTGAAAGAGGAGGACTGCACCTGTGTGGTACTTAT--- 921
 QY 293 GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSer 312
 Db 922 ---GGTTCATCTCATCCATATGCGGATGTAGGAATTAGCAACGATTAATCTCGATCTT 978
 QY 313 TyrIleGlyArgAsnProGluAsnAspAsnPhetheThrGlu-----SerPheSer 330
 Db 979 TATATGGGAACAAACCTGCCANTGCAATTTTACCTTTGCGGATGCGAATTTCACTAAG 1038
 QY 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
 Db 1039 CCACCT-----TCAAGAGTTTACAAACAGCGCTGATGCGAGATCTTTGTTTCATTTTGG 1089
 QY 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
 Db 1090 GAAAGTACCGAAACAGCAGCAGAGGTTTCAGCAAGAAAAACAGAGCTCAGAACCAAGTA 1149
 QY 371 LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu 390
 Db 1150 CTTGAAGCCATGCTCACAGACTTTCATATTCACAAATAGCTGATAGTCTCGGAAAAATC 1209
 QY 391 SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro 410
 Db 1210 TTGTTTGGCATTTTCAGAGAGTCTTCAAGTGTCTTAAACAAAGTACGCTCTGCTGGCAACCT 1269
 QY 411 LeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430
 Db 1270 CTAGTCGATGACTGCAACTGCCCTTAAATAATCAGGTGAGAGCTTTTCGAGAGGCACTGTGA 1329
 QY 431 AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly 450
 Db 1330 TCGCTCTCTCAGTACGGTATCAAGCACATGAGTCTTTTGGCAACATCTGCAATCCAGGG 1389
 QY 451 ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer 465
 Db 1390 ATTCAATGAGCAATTCGAGGAGGAGGCTTCACAGGCTTGTGTACC 1434

RESULT 14

US-09-708-427-10759
 ; Sequence 10759, Application US/09708427
 ; GENERAL INFORMATION:
 ; APPLICANT: N. ALEXANDROV et al.
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
 ; FILE OF INVENTION: THEREBY
 ; FILE REFERENCE: 2750-1243P
 ; CURRENT APPLICATION NUMBER: US/09/708,427
 ; CURRENT FILING DATE: 2000-11-09
 ; NUMBER OF SEQ ID NOS: 85364
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 10759
 ; LENGTH: 1485
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana

FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..1485
OTHER INFORMATION: any n = a, g, c, t, unknown, or other
NAME/KEY: misc_feature
LOCATION: 1..1485
OTHER INFORMATION: Ceres Seq. ID 1821917
US-09-708-427-10759

Alignment Scores:
Pred. No.: 1.64e-113 Length: 1485
Score: 1246.00 Matches: 233
Percent Similarity: 69.89% Conservative: 71
Best Local Similarity: 53.56% Mismatches: 121
Query Match: 50.77% Indels: 10
DB: 28 Gaps: 4

US-09-934-066-2 (1-466) x US-09-708-427-10759 (1-1485)

Qy 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
Db 145 GAAACGACGACGATTAACCTCCGCTACTAGTGGCTGTTCTAGTCGCGGATCAGC 204
Qy 56 GluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
Db 205 GGATATTGGAATTACAGGCATCAGGCTGATATATGCCATCGCTTATCAACTTCTGAGGAA 264
Qy 76 GlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSer 95
Db 265 GTGGGATGGAAGAGAGAGATATTGGTATTCATGATGATGATATTCGTACAAATTAC 324
Qy 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGly 115
Db 325 GAGATCCAAAGGCCTGGAACCATTAACACAGCCCTCATGGAAGAGATGCTTATCAAGGA 384
Qy 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLys 135
Db 385 GTTCCCAAGGATTAATCTGAGATGATGTCGAATGTTGATAATCTATTGCTGTGATCCCT 444
Qy 136 GlyAsnGluSerGlyValThrGlyValAsnGlyLysValValLysSerGlyProAsnAsp 155
Db 445 GGAGACAAACTCCTGTTAAAGGGGAGGAGTGGAGGTTGTGGATAGTGGCTCATATGAT 504
Qy 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
Db 505 CATATCTTCATATCTACAGTACCACCTGGTGGTCTGGAGTCTTGGGATGCCAACTTCT 564
Qy 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
Db 565 CTTTACCTATATGCAATGATGTCATGATGTCGGAAGAGAAACATGCTTTAGGAACA 624
Qy 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIle 215
Db 625 TATAAAGCTGTGTTTATCTCGAAGCTTGGCATCTGGAAGTATCTTTGAAGGCTT 684
Qy 216 LeuLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp 235
Db 685 CTTCTCGAGGTTTGAACATCTATGCCACAACTGATCAACCCGCGGAGGAGCAGTTGG 744
Qy 236 GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGly 255
Db 745 GGACCTATTGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 804
Qy 256 AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
Db 805 GACTTGTACAGTCTGCTGGATGGAGATAGTGGTATGCACAAATTACAGACTGAGACT 864
Qy 276 LeuGluGlnGlnTyrHisValValLysArgArg-----ValGlySerAspValPro 292
Db 865 CTCACACGAGCAATGAACTTGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 921
Qy 293 GluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSer 312
Db 921 GATATTTGGAATTACAGGATGAGGCTGATATATGATGATGATGATGATGATGATGATGATGAT 364

Db 922 ---GGTTCTCATGTCATGCAATATGCGGATGTAGGAATTAGCAAGGATATATCTCGATCTT 978
Qy 313 TyrIleGlyArgAsnProGluAsnAspAsnPheThrGlu-----SerPheSer 330
Db 979 TATATGGGAACAACCCCTGCCAATGACAATTTACCTTTGCGGATGCGAATTCACATAAG 1038
Qy 331 SerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGln 350
Db 1039 CCACCT-----TCAAGAGTTACAACACGCGTGCATGTCAGATCTGTGTTCATTTTGG 1089
Qy 351 ArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 370
Db 1090 GAAAGTACCGAAAGCACCAGAGGTTCCAGCAAGAAACACAGAGCTCAGAAGCAAGTA 1149
Qy 371 LeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeu 390
Db 1150 CTTGAAGCCATGCTCACAGACTTCATATTGACAATAGCGTGATCTCGTCGCAAAATC 1209
Qy 391 SerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnPro 410
Db 1210 TTGTTTGGCATTTTCGAGAGGTCCTGAAGTCTTAACAAAGTACGCTGCTGGCAACCT 1269
Qy 411 LeuValAspAspThrAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGly 430
Db 1270 CTAGTCGATGACTGGAACTCCCTTAAATAATCAGGTGAGAGCTTTCGAGAGGCACTGTGGA 1329
Qy 431 AlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGly 450
Db 1330 TCCTGCTCTCAGTACGCTATCAGCACACATGAGTCTTTTGCACAACTCTCAGTGCAGG 1389
Qy 451 ValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSer 465
Db 1390 ATTCAATGGAGCAATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434

RESULT 15

US-09-935-625-11673
Sequence 11673, Application US/09935625

GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
FILE REFERENCE: 2750-1481P
CURRENT APPLICATION NUMBER: US/09/935, 625
CURRENT FILING DATE: 2001-08-24
NUMBER OF SEQ ID NOS: 33136
SEQ ID NO 11673
LENGTH: 1485
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: Misc_feature
LOCATION: 1..1485
OTHER INFORMATION: Ceres Seq. ID no. 1821917
US-09-935-625-11673

Alignment Scores:
Pred. No.: 1.64e-113 Length: 1485
Score: 1246.00 Matches: 233
Percent Similarity: 69.89% Conservative: 71
Best Local Similarity: 53.56% Mismatches: 121
Query Match: 50.77% Indels: 10
DB: 28 Gaps: 4

US-09-934-066-2 (1-466) x US-09-935-625-11673 (1-1485)

Qy 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsn 55
Db 145 GAAACGACGACGATTAACCTCCGCTACTAGTGGCTGTTCTAGTCGCGGATCAGC 204
Qy 56 GluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75
Db 205 GGATATTGGAATTACAGGATGAGGCTGATATATGATGATGATGATGATGATGATGATGATGAT 364

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2003, 18:27:24 ; Search time 691 Seconds
(without alignments)
3732.872 Million cell updates/sec

Title: US-09-934-066-2

Perfect score: 2454

Sequence: 1 MSPLGHQILVFLHLLIF.....CNMGVDVKQTVSAIEQACSM 466

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 6653949 seqs, 2767611981 residues

Total number of hits satisfying chosen parameters: 13307898

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-O/cgn2_1/USPTO.spool/US0934066/runat_19052003_163701_24433/app_query.fasta_1.647
-DB=Pending Patents.NA.New -ORFMT=fastap -SUFFIX=p2n.rnnp -MINMATCH=0.1
-LOOPCI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR=SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0934066.rcgn_1_1_327 @runat_19052003_163701_24433
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAL -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPEXT=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2.*
7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3.*
8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2.*
10: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287.5	52.4	1758	8	US-10-425-114-8500
2	1284.5	52.3	1937	8	US-10-424-599-76825
3	1279.5	52.1	2133	8	US-10-424-599-59168
4	1273	51.9	2006	8	US-10-424-599-59165
5	1205.5	49.1	1881	8	US-10-425-114-28946
6	1204.5	49.1	1786	8	US-10-425-114-35367
7	1204.5	49.1	1794	8	US-10-425-114-30424

ALIGNMENTS

RESULT 1

US-10-425-114-8500

; Sequence 8500, Application US/10425114

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; SEQ ID NO 8500

; LENGTH: 1758

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700756373_FLI

US-10-425-114-8500

Alignment Scores:

Pred. No.: 3.27e-121

Score: 1287.00

Percent Similarity: 68.82%

Best Local Similarity: 52.47%

Query Match: 52.44%

DB: 8

Length: 1758

Matches: 244

Conservative: 76

Mismatches: 135

Indels: 10

Gaps: 4

Sequence 28422, A
Sequence 15845, A
Sequence 15458, A
Sequence 34046, A
Sequence 27715, A
Sequence 35501, A
Sequence 49438, A
Sequence 10538, A
Sequence 49416, A
Sequence 19102, A
Sequence 6839, Ap
Sequence 15848, A
Sequence 30, Appl
Sequence 21292, A
Sequence 452, App
Sequence 3, Appl
Sequence 6167, Ap
Sequence 33, Appl
Sequence 1847, Ap
Sequence 4031, Ap
Sequence 4031, Ap
Sequence 114, App
Sequence 32083, A
Sequence 32083, A
Sequence 32078, A
Sequence 32078, A
Sequence 32084, A
Sequence 32084, A
Sequence 32079, A
Sequence 32079, A
Sequence 71, Appl
Sequence 446, App
Sequence 9493, Ap
Sequence 30081, A
Sequence 38671, A

Qy	2	SerSerPro---	LeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe	20
Db	148	TCCACACCATCATCGGACCGTTTTTCGGATCCTCTTCTCCTCGCCACCCCTCATCACCTC	207	
Qy	21	SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGlu	36	
Db	208	GCCTCCGGTCCGCCACGATATTCTCCGGTTACCTCCGAAAGCATCCACTTTTTTCAAA	267	
Qy	37	-----SerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGly	53	
Db	268	GCACCTGCTAATCCGATCAAAACGATGAGGGACCCAGCTGGCGCGTTTACTTCCCGT	327	
Qy	54	SerAsnGluTyrTyrrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeu	73	
Db	328	TCCAATGCTACTCGAATTCACAGSCACCACTCTGATTTGGCATGTCATATCAACTACTG	387	

Qy	74	ArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPhe	93
Db	388	AGAAAGGTCGTGAAAGAGGAAAATATTGTCGTATTATGTATGATGACATTGCTTTC	447
Qy	94	SerSerGluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyr	113
Db	448	ANTGAAGAACCCACGCGCTGGAGTCATTATTAAACAGCTCCACACGGAATGATGTTTAC	507
Qy	114	LysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnVal	133
Db	508	AAGGAGCTTCTTAAGGATTAACGTTGGTGAAGATCTTACTGTTGACAACTTTTTCGCTGCT	567
Qy	134	LeuLeuGlyAsnGluSerGlyValThrGlyAsnGlyLysValLysSerGlyPro	153
Db	568	ATACTTGGAAATAGTCAGCTCTTACTGGTGGCAGTGGGAAGGTTGTGATAGTAGTGGCCCC	627
Qy	154	AsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetPro	173
Db	628	ANTGATCATATATTATATACTACTCTGATCATGCGCGTCCGGAGTGCTAGGAGTCCT	687
Qy	174	ThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArg	193
Db	688	ACTAATCCATACATGATGATCGATCCGATCTGATTGAAGTCTTGAAGAAGAAGCATCTCT	747
Qy	194	LysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGlu	213
Db	748	GGAACTTTAAAAGCCTAGTATTATTCTAGAGGCATGTGAATCTGGAGATATCTTTGAA	807
Qy	214	GlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSer	233
Db	808	GGTCTTCTCCAGAAGTCTGAATATCTATGCAACACAGCTTCAANTGCTGAAGAAAGC	867
Qy	234	SerTrpGlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCys	253
Db	868	AGTTGGGAACATATTCTCTCGGGAGATCTCTAGTCTCCCTGGAATATGAAACCTGC	927
Qy	254	LeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLys	273
Db	928	CTGGTGACCTGTACAGTGTTGCTTGATGGAAGATAGTACATACACAATTCGCGAACA	987
Qy	274	GluThrLeuGluGlnTyrHisValLysArgArgVal-----GlySerAspVal	291
Db	988	GAAACTTTACATCAACAATACGACTTGGTCAAAAGAGGACTATGANTGAAATTCATC	1047
Qy	292	ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer	311
Db	1048	---TATGGTTCCCAAGTCGATCGCATGTGTCAGATAGGTCATAGGCGTTAGCAAGAACA	1104
Qy	312	SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSer	331
Db	1105	TTATATTGGGTACAAATCTCGTAAATGATAATTACTTTTGGCATAAA---AACTCA	1161
Qy	332	ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArg	351
Db	1162	TTGGTGCCACCTTCAAAAGCAGTCAACCAAGTCATGCAGATCTCATTCCTCGGAT	1221
Qy	352	LysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeu	371
Db	1222	AAGTTCGCAAAAGCTCTCTGGGTTCTTCAGGAAAGCTGCAGCTGAGAAGAAATTCG	1281
Qy	372	AspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSer	391
Db	1282	GAAGCAATGTCTCACAGAATGCATATAGATGACAAACATGAACATTATTTGGAAGCTCTTA	1341
Qy	392	ValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlnProLeu	411
Db	1342	TTTGGCATTTCAAAGGGTCCAGAAGCTCTTAGCAGTGTAGACCTGCTGGCAACCACTT	1401
Qy	412	ValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAla	431
Db	1402	GTGTGACGTGGAGCTGCCCTTAAACACTGGTTAGGACTTTTTCAGACACATTTGGTGA	1461

Qy 432 ThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyVal 451
Db 1462 CTGCTCAGTATGGGATGAACATATGAGTCTCTTGCACAACTTCTGCAACGCTGGAATA 1521
Qy 452 AspValLysGlnThrValSerAlaIleGluGlnAlaCys 464
Db 1522 CGGAACAGCAAAATGGCTGAGGCTCGGCACAGCATGT 1560

RESULT 3
US-10-424-599-59168
: Sequence 59168, Application US/10424599
: GENERAL INFORMATION:
: APPLICANT: La Rosa Thomas J
: APPLICANT: Kovalic David K
: APPLICANT: Zhou Yihua
: APPLICANT: Cao Yongwei
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53223)B
: CURRENT APPLICATION NUMBER: US/10/424,599
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 59168
: LENGTH: 2133
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT3847_24440C.1
US-10-424-599-59168

Alignment Scores:
Pred. No.: 2,52e-120 Length: 2133
Score: 1279.50 Matches: 245
Percent Similarity: 68.40% Conservative: 71
Best Local Similarity: 53.03% Mismatches: 135
Query Match: 52.14% Indels: 11
DB: 4 Gaps: 4

US-09-934-066-2 (1-466) x US-10-424-599-59168 (1-2133)

Qy 13 PheLeuHisAlaLeuLeuIlePhe-----SerAlaGluSerArgLysThr 27
Db 28 TTCTCTCCACGGCTCCCTCCCTTCATAGCCTTCGCCACCTCTGCTCTCCGGCCGCCGT 87
Qy 28 GlnLeuLeuAsnAsp-----AsnAspValGluSerSerAspLysSerAlaLys 43
Db 88 GACCTCGCTCGGAGACTTCTCTCGGCTCCCTCCGAAACTGATAACGACGACACTTCAAG 147
Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGln 63
Db 148 GGCACCCGGTGGGCGCTCCCTCCCTCGGCGGTCCCAATGGTTACTTGGAAATACAGACATCAG 207
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
Db 208 GCTGATGTTGTTCACGCCATCAAAATATTAGGAAAGGTGGTCTGAAAGAAAGAAATATT 267
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 268 ATTCGTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 327
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 328 ATTAACAAACCCAGATGGAGGTGATGTTTATAAAGGAGTTCCAAAGGATTACACCGCGAA 387
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 388 GATGTTACTCTGTATCACTACTTTTTTGTCTGCTTTACTTTGAAATTAAGTCAGCACTGAT 447
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 448 GGCAGTGGGAAGGTTGGACAGTGGTCCCTGATGATCATATATTTGATATCTACTACTGAC 507
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183

```
Db 508 CATGAGGTCCTGGGCTCTGGGATGCTCTGCTGCTTACTTATACGGCGGATGATCTG 567
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 568 ATTGAAGTCCTTGAAGAAACATGCTCTCTGGAACATATAAAACCTAGTATTTATCTG 627
Qy 204 GluAlaCysGluSerGlySerMetPheGluGlyLleLeuLysLysAsnLeuAsnIleTyr 223
Db 628 GAGGATGTCGAATCTGGAGATCTCTTGAAGTCTCTCTGGAAGATATCAATATTTAT 687
Qy 224 AlaValThrAlaAlaAsnSerLysGluSerSerTyrPheGlyValTyrCysProGluSerTyr 243
Db 688 GCAACCACTGCTTCAATGCAAGAAAGTATGTTGGGAACATATTCGCCCGGGAGTAT 747
Qy 244 ProProProProSerGluIleGlyLysCysLeuGlyAspThrPheSerIleSerTyrLeu 263
Db 748 CCTAGTCTCTCCCAAGATATACAACTGTTTGGGTGACTTGTACACTGTTGCTGGATG 807
Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHisValVal 283
Db 808 GAAGACAGTGACAGACAAATTTGGAACACAACTCTGCACCAACATATAAATGGTT 867
Qy 284 LysArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThr 302
Db 868 AAGACAGCATATATCTGGAGATTCATACATATGGCTCTACGGTATGATGATGATGAT 927
Qy 303 GluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsn 322
Db 928 GTAGGCTTAGCAGAGATGTTCTCTCCATATTTGGGTACAGATCTGCTGAATGATAAT 987
Qy 323 PheThrPheThrGluSerPheSerSerProLysSerAsnSerGlyLeuValAsnProArg 342
Db 988 TTCACCTTTTGGGATGAA---AACTCCTTATGGTCACTTCAAAACCAAGTCAACCAAGT 1044
Qy 343 AspIleProLeuLeuTyrLeuGluArgLysIleGlnLysAlaProMetGlySerLeuGlu 362
Db 1045 GATGCTGATCTCATCTCTTGGGATAAGTTCGCAAGCTCTGAGGCTCTCTCAGG 1104
Qy 363 SerLysGluAlaGlnLysLysLeuAspGluLysAsnHisArgLysGlnIleAspGln 382
Db 1105 AAAATACAGCTTCAGAAACAAAGTTTGGAGCAATGTCTCACAGAAATGCGATGAGACAAC 1164
Qy 383 SerIleThrAspIleLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThr 402
Db 1165 AGTGTAAACTGATTGGGAAGCTTTTATTGGCATTGAAAAGGTCACGAAGTACTCAAC 1224
Qy 403 SerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThrLeuVal 422
Db 1225 GCTGTTAGACGGCTGGATCGCACTTCTTCTGATGCTGCACTGCTGCAAAACCATGGTG 1284
Qy 423 AsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAla 442
Db 1285 AGGACTTTTGACACATTTGGATCCTGTCTCAATACGGGATGGAACACATGAGTCC 1344
Qy 443 LeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluGln 462
Db 1345 TTTGCAACATCTGCAATGTAGGATAAAGAAATGAACAAATGGCTGAGGCTTCAGCACAA 1404
Qy 463 AlaCys 464
Db 1405 GCTTGT 1410
```

RESULT 4

```
US-10-424-599-59165
; Sequence 59165, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: plants and Uses Thereof for Plant Improvement
```

```
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424-599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 59165
; LENGTH: 2006
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_24438C.1
US-10-424-599-59165
```

```
Alignment Scores:
Pred. No.: 1,07e-119 Length: 2006
Score: 1273.00 Matches: 236
Percent Similarity: 71.03% Conservative: 68
Best Local Similarity: 55.14% Mismatches: 122
Query Match: 51.87% Indels: 2
DB: Gaps: 2
```

US-09-934-066-2 (1-466) x US-10-424-599-59165 (1-2006)

```
Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 122 ACGATGACACACGTCAGGCGCCGCTGCGCTCTCTCGCGGTTCCATGCTAC 181
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77
Db 182 TGGAAATACACACATCAGGCTGATGTTGTGACGCCCTATCAATACTTAAGAAAGTGGC 241
Qy 78 LeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn 97
Db 242 CTGAAGAAGAAATATATTTATTTATGATGATGATGATGATGATGATGATGATGATGAT 301
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 302 CCAAGGCTGAGGATCATCATTAACAAACAGATGAGGATGATGATGATGATGATGATGATGAT 361
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 362 AAGGATTACACCGCGAAGATGTTACTGTTGGTAACCTTTTCTGCTGCTTACTTGGAAAC 421
Qy 138 GluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIle 157
Db 422 AAGTCGGCATTACTGTCGACAGTGGCAAGCTTGTGACAGTGGACCTGATGATCATATA 481
Qy 158 PheIleTyrTrpAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 482 TTTGTTATATATCTAGCATGAGGTCAGGTCAGGTCGCGGATGCTGCTGCTGCTTAC 541
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsn 197
Db 542 TTATATCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 601
Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys 217
Db 602 AACCTAGTATTTATCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 661
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 662 GAAGATATCAATATTTATGCAACCAACAGTTCGAATGCAAGAAAGATAGTTGGGAACA 721
Qy 238 TyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThr 257
Db 722 TATTGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781
Qy 258 PheSerIleSerTrpLeuGluAspSerAspLeuIleAspMetSerLysGluThrLeuGlu 277
Db 782 TACAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
Qy 278 GlnGlnTyrHisValValLysArgArg---ValGlySerAspValProGluThrSerHis 296
Db 842 CAACAATATAAATTTGGTTAAAGACAGCATTATATCTGGACATTCATATGCTGCTCTCAC 901
```

Qy	297	ValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArg	316
		: : : : : : : : : : : : : : : :	
Db	902	GTGATCAGTATGGTGTATGACGGCTAGCAGTGTCTCTCTCCATTATTTGGGTACA	961
Qy	317	AsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSer	336
Db	962	GATCCTTGCTAATGATAAATTTTCTTTTGTGGAGCAA---AACTCCTTATGGTCACCTTCA	1018
Qy	337	GlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAla	356
Db	1019	AAACCAAGTCAACCAACGTGATGCTGATCTCATCCATTTTGGGATAGTTCCGCAAGCT	1078
Qy	357	ProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHis	376
Db	1079	CCTGAGGGTCTCTCAGGAAAAATGCAAGCTCAGAACAAAGTTTGGAAAGCAATGCTCTCAT	1138
Qy	377	ArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsn	396
Db	1139	AGAATGCATGATAGCAACAGTGTAAACCTATTGGGAAGCTTTATTTGGTATTGAAAAG	1198
Qy	397	ValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTTPasp	416
Db	1199	GGTCCAGAAGTACTCAACGCTGTATGACCGCAGGATCGGCACCTGTTGATGATGGCAC	1258
Qy	417	CysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGly	436
Db	1259	TGCCTTAAAAACCATGCTGAGGACTTTTGAGACACATTTGGATCCTTGCTCAGTATGGC	1318
Qy	437	LeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThr	456
Db	1319	ATGAACATATGATAGGTCCTTTGCAACACTCTGCAATGTAGGGATAAGAAATGAGCAATG	1378
Qy	457	ValSerAlaIleGluGlnAlaCys	464
Db	1379	GCTGAGGCCTCTCCACAGCTTGT	1402
RESULT 5			
US-10-425-114-28946			
; Sequence 28946, Application US/10425114			
; GENERAL INFORMATION:			
; APPLICANT: Liu, Jingdong			
; APPLICANT: Zhou, Yihua			
; APPLICANT: Kovalic, David K.			
; APPLICANT: Screen, Steven E.			
; APPLICANT: Tabaska, Jack E.			
; APPLICANT: Cao, Yongwei			
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With			
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement			
; FILE REFERENCE: 38-21(53313)B			
; CURRENT APPLICATION NUMBER: US/10/425,114			
; CURRENT FILING DATE: 2003-04-28			
; NUMBER OF SEQ ID NOS: 73128			
; SEQ ID NO 28946			
; LENGTH: 1881			
; TYPE: DNA			
; ORGANISM: zea mays			
; FEATURE:			
; OTHER INFORMATION: Clone ID: LIB4767-010-E5_FLI			
US-10-425-114-28946			
Alignment Scores:			
Pred. No.: 7,96E-113 Length: 1881			
Score: 1205.50 Matches: 227			
Percent Similarity: 68.66% Conservative: 71			
Best Local Similarity: 52.30% Mismatches: 129			
Query Match: 49.12% Indels: 7			
DB: 8 Gaps: 3			
US-09-934-066-2 (1-466) x US-10-425-114-28946 (1-1881)			
Qy	36	GluSerSerAspLysSerAlaTyrGlyThrArgTTPAlaValLeuValAlaGlySerAsn	55
		::: : : :	

207	Db	GAGCAACGGATGACGAGCGCGTGGGACCGCGTGGCGTGCTGCTCATCGCGCGCTCCAAC	266
56	Qy	GLuTYrTYrAsnTYrArgHisGlnAlaAspIleCYsHisAlaTYrGlnIleLeuArgLys	75
267	Db	GGCTACTACAACATACCGCCACACAGCGGACATCTGCCATGACATACCCAGATCATGAAGAAG	326
76	Qy	GlyCYLeuLysAspGluAsnIleIleValPheMetTYrAspAspIleAlaPheSer	95
327	Db	GGCGGACATTAAAGACGAGAAACATCGCTGTCTTCATGTACGATGACATCGCGCATACGCCG	386
96	Qy	GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTYrLysGly	115
387	Db	GAATAATCCGAGGCGTGTGTTCATCAATAATCATCCAGGTCGCCAGCGTCTATGCTGGG	446
116	Qy	ValProLysAspTYrThrLysGluAlaValAsnValGlnAsnPheTYrAsnValLeuLeu	135
447	Db	GTGCCAAAGATTACACTGGCGCAGAGGTCAACGTCGACAAATTTCTTCGCTGTTCTGCTT	506
136	Qy	GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValLysSerGlyProAsnAsp	155
507	Db	GGCAACAAAACACTCTCTCAGGGGTGGGAGCGGCAAGGTTGTGGACATAGTGCCCAATGAT	566
156	Qy	AsnIlePheIleTYrTYrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly	175
567	Db	CATATATTTTGTTCATACAGTACCATGGGGTCTGGCGTCTTGGAAATGCTCATGCTAT	626
176	Qy	AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMethLysArgLysLys	195
627	Db	CCATATCTCTATGTTGATGACCTCGTAGATGTCTCTGAAGAAGAAGCATGCTGCCGGGACC	686
196	Qy	TYrAsnLysMetValIleTYrValGluAlaCYsGluSerGlySerMetPheGluGlyIle	215
687	Db	TACAAAAGCCTGCTCTTTACCTTGAAGCATGCGAATCTGGGAGCATCTTTGAGGCGCTC	746
216	Qy	LeuLysLysAsnLeuAsnIleTYrAlaValAlaAlaAsnSerLysGluSerSerTrp	235
747	Db	CTCGCGGAATGACATAAATGTATGCGCACCCGCGTCAATGACAGAGAGATGACTGG	806
236	Qy	GlyValTYrCysProGluSerTYrProProProSerGluIleGlyTYrCysLeuGly	255
807	Db	GGGACGTACTGCCCTGGCGAGTTCGCCAGCCCTCCACCGAGTATGACACTTCTCTGGGA	866
256	Qy	AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr	275
867	Db	GACCTGTATAGTGTGTTGCATGGGAAGACAGTGATTTCCACAATCTGCCAACTGAATCT	926
276	Qy	LeuCluGlnTYrHisValValLysArgValGlySerAspValProGluThr---	294
927	Db	CTCAAGCGAAATACAACTTTGGTCAAGATTAGGACAGCG-----GTTACAGGATACATTC	980
295	Qy	-----SerHisValCYsArgPheGlyThrGluLysMetLeuLysAspTYrLeuSer	311
981	Db	AGCTATGGCTCCCATGTGATGCAATATGGTTTCATTGGAGTTGAATGTTAAGCATCTGTTT	1040
312	Qy	SerTYrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSer	331
1041	Db	TCGTACATTGGCACAAACCTCGTTAACGATGACAAACAGCTTATAGAAGACAACTCGTTG	1100
332	Qy	ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTYrLeuGlnArg	351
1101	Db	CCA---TCATTCTCAAAAGCGTGTAAATACGCGCGAGCGTGACCTGTCTACTTCTCGCAG	1157
352	Qy	LysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeu	371
1158	Db	AAGTACCGGAATTTGGCAGACACTCATATGACAGAAATGAAGCTCGGAGGGAGTTGCTT	1217
372	Qy	AspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSer	391
1218	Db	GAAGTGTATGCCACACAGTCTCATGTGTGACAGCAGTGTTGAGCTCATTTGGACGCTTCTC	1277
392	Qy	ValLysGlnThrAsnValLeuAsnLeuLeuTYrSerThrArgThrThrGlyGlnProLeu	411
1278	Db	TTTGGCTCTCAGAGACGGTCCAAGGTTCTCAAAAGCGTCTCGCGAGCTGGTGAGCTCTC	1337

Qy 282 valvalLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
:::||||| ||| ||| ::::| |

US-09-934-066-2 (1-456) x US-10-425-114-28422 (1-1817)

Qy 42 AlaLysGlyThrArgTrrAlaValIeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61

```
Db 215 GCGTCGGGACCGGTCGGCGCTGCTCATCGCGGCTCCACGGCTACTACAACTACCGC 274
Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
Db 275 CACCAGCGGACATCTGCCATCACCAGATCATCAAGAAGGCGGACTTAAAGACGAG 334
Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 335 AACATCGTTGCTTCATGTACCATCGGCATACGCCGGAATCCGAGCGCTGGT 394
Qy 102 ValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 395 GTCATCATTAATCATCCAGGTCGGCAGCTCTATGCTGGGTGCCAAGGATTAACAT 454
Qy 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 455 GGGCGAGAGGTCACGTCGACAAATCTTCGCTGTTCGTGGCAACAACTGCTCTC 514
Qy 142 ThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 515 AGGGTGGGAGCGGCAAGGTTGTGCACAGCTGGCCCAATGACCATATATTTGTTTCTAC 574
Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
Db 575 AGTGACATGGGGCTCGGCTGCTGGAATGCTGCTATCCATATCTCTATGGTGAT 634
Qy 182 AspPheAsnGluValLeuGluLysMethHisLysArgLysLysTyrAsnLysMetValIle 201
Db 635 GACCTCGTAGATGCTCTGAAGAAGACGATGCTGCCGGGACCTACAAAAGCGCTGCTTT 694
Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 695 TATCTTGAAGCATGCGAATCTGGGAGCATCTTTGAGGCGCTCTGCGCAATGACATAAT 754
Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTyrGlyValTyrCysProGlu 241
Db 755 GTGTATGCGACCCCGCTCAATCAGAGGAGAGTACCTGGGGGACCTACTGCGCTGGC 814
Qy 242 SerTyrProProProSerSerLueGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db 815 GAGTTCGGCGGCTCCCGGAGTATGACACTGCTGTGGAGACCTGTATAGTGTGCT 874
Qy 262 TrpLeuGluAsnSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHis 281
Db 875 TGGATGGAAGACGATGTTCCCAATCTGCAACTGCACTGAATCTCAAGCAGCAATACAC 934
Qy 282 ValValLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
Db 935 TTGGTCAAGGATAGGACAGCG-----GTTCAGGATACATTCAAGCTATGCTCCCATGTG 988
Qy 298 CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
Db 989 ATGCAATATGTTTCATTTGGAGTTGAATGTTAAGCATCTGTTTTCGTATCTGGCACAAC 1048
Qy 318 ProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGly 337
Db 1049 CTGCTTAACGATGACAAACGTTTATAGAACAACACTCGTTGCCA---TCGTTCTCAAG 1105
Qy 338 LeuValAsnProArgAspIleProLeuGlnArgLysIleGlnLysAlaPro 357
Db 1106 GCTGTTAATCAGCGGACGCTGACCTTCTTCTTCTGGCAGAAAGTACCGGAAATTCGCA 1165
Qy 358 MetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
Db 1166 GACAGCTACCTGAGAAAAATGAAGCTCGGAAGGAGTGTCTGAAGTCATGCGCCACAGG 1225
Qy 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValGlnThrAsnVal 397
Db 1226 TCTCATGTTGACAGAGTGTGACGTCATGGAAGCCCTTCTCTTGGCTCTGAGGACGGT 1285
Qy 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspTyrTrpAspCys 417
```

```
Db 1286 CCAAGGTTTCTGAAGCCGCTCCGTGCAGCTGTGTGAGCCTCTGTGTCGATGATTGGAGCTGT 1345
Qy 418 PheLeuValLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeu 437
Db 1346 CTCAAGTCCACGGTTCGTACTTTTGGAGCGCAATCTGGTCTGTTGGCCGCACTATGGGATG 1405
Qy 438 LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrVal 457
Db 1406 AAGCAGATCGGTCCTTCGCAACATCTGCAACGCTGGCATCCTTCTTCCGAAGCAGTGTG 1465
Qy 458 SerAlaIleGluGlnAlaCysSer 465
Db 1466 AAGTCGCTGCTCAGCGCTTGCACC 1489

RESULT 9
US-10-425-114-15845
; Sequence 15845, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15845
; LENGTH: 1769
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-088-P9_FLI
US-10-425-114-15845

Alignment Scores:
Pred. No.: 1 878-112 Length: 1769
Score: 1201.50 Matches: 226
Percent Similarity: 69.16% Conservative: 70
Best Local Similarity: 52.80% Mismatches: 125
Query Match: 48.96% Indels: 7
DB: 8 Gaps: 3

US-09-934-066-2 (1-466) x US-10-425-114-15845 (1-1769)
Qy 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 191 GCGTCGGGACCGGTCGGCGCTGCTCATCGCGGCTCCAAACGGCTACTACAACTACCGC 250
Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
Db 251 CACCAGCGGACATCTGCCATCAGATCAGATCATGAAGAAGGCGGACTTAAAGACGAG 310
Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 311 AACATCGTTGCTTCATGTACCATCGCATCGCGCATACCGGAAATCCGAGCGCTGGT 370
Qy 102 ValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 371 GTCATCATTAATCATCCAGGTCGGCAGCTCTATGCTGGGTGCCAAGGATTAACAT 430
Qy 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 431 GGGCGAGGCTCAGCTGCAACATTTCTCGCTGTCTTCTGGCAACAACTGCTCTC 490
Qy 142 ThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 491 AGGGTGGGAGCGGCAAGGTTGTGCACATGCGCCCAATGACCATATATTTGTTTCTAC 550
Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
```

```

Db 551 AGTGACCATGGGGTCTCGCTGCTTGAAGTCCCTACGATATCCATCATCTCTATGCTGAT 610
QY 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysLysLysAsnLysMetValIle 201
Db 611 GACCTCGTAGATCTCTGAAGAGAACGAGCTGCGGGACCTACAAAAGCCTGGTCTTT 670
QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 671 TACCTTGAAGCATGCCAATCTGGAGCATCTTTGAGGGCTCTCTGCCGAATGACATAAAT 730
QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db 731 GTGATGCGACCAACCGCTCAATGAGAGAGAGTAGCTGGGGACGCTACTGCCCTGCG 790
QY 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db 791 GAGTCCCGAGCCCTCCAGGAGTAGACACTGCTTGGGAGACCTGATAGTGTGCT 850
QY 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnIntyrHis 281
Db 851 TGGATGGAAGACAGTGATTTCCACAATCTGCGAATCTCTCAAGCAGCAATACAAC 910
QY 282 ValValLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
Db 911 TTGGTCAAGGATAGGACACG-----GTTCAGGATACATTCAGCTATGGCTCCCATGTC 964
QY 298 CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
Db 965 ATGCAATATGGTTCATGGAGGTGAATGTAAGCATCTGTTTCGTACATTTGGCACAAAC 1024
QY 318 ProGluAsnAspAsnPheThrPheThrGluSerPheSerSerProIleSerAsnSerGly 337
Db 1025 CCTGTACGATGACACACAGTTTCATAGAGACAACTCGTTGCCA---TCATCTCAAG 1081
QY 338 LeuValAsnProArgAspIleProLeuTyrLeuGlnArgLysIleGlnLysAlaPro 357
Db 1082 GCTGTAAATCAGCGGACGCTGACCTGTGTCTTCTTGGCAGAAGTACCAGAAATTGGCA 1141
QY 358 MetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
Db 1142 GACAGCTCACATGAGAAATGAAGCTCGGAGGAGTGTCTTGAAGTATGGCCACAGG 1201
QY 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
Db 1202 TCTCATGTTGACAGCATGTGAGCTCATTTGGAAGCCTTCTCTTGGCTCTGAGGACGCT 1261
QY 398 LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCys 417
Db 1262 CCAGGGTTTCTGAAGCCGCTCGCTGAGCTGTGAGCCTCTCGTCATGATTGGAGCTGT 1321
QY 418 PheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeu 437
Db 1322 CTCAGTCCAGGTTCTGCTACTTTTGAAGCGCAATGTGGTGTGGCGAGTATGGGATG 1381
QY 438 LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrVal 457
Db 1382 AAGCACATGGGTCCTTCGCAAACTCTGCAAGGCTGGCATCTCTCTCAAGCAGTGTGCG 1441
QY 458 SerAlaIleGluGlnAlaCysSer 465
Db 1442 AAGGTCGCTGCTCAGGCTTGCACC 1465

```

RESULT 10

US-10-425-114-15458

; Sequence 15458, Application US/10425114

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

```

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 15458
; LENGTH: 1774
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-003-F12_FLI
US-10-425-114-15458

```

Alignment Scores:

```

Pred. No.: 1-88e-112 Length: 1774
Score: 1201.50 Matches: 226
Percent Similarity: 69.16% Conservative: 70
Best Local Similarity: 52.80% Mismatches: 125
Query Match: 48.96% Indels: 7
DB: 8 Gaps: 3

```

US-09-934-066-2 (1-466) x US-10-425-114-15458 (1-1774)

```

QY 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
Db 196 GCCGTGGGACCGCGTGGCGTGCATCGCGGCTCCAAACGGCTACTACAACACTACCGC 255
QY 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGlu 81
Db 256 CACCAGCGGACATCTGCCATCATACCAGATCATGAAGAAGCGGAGCTTAAGAGCAG 315
QY 82 AsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
Db 316 ACATCGTGTCTTCTATGATGACATGACATGCGCATAGCCCGGAAATCCGAGGCGCTGT 375
QY 102 ValIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
Db 376 GTCATATAATCATCCCGAGGTGGCGAGCTATGCTGGGTGCCAAGGATTACACT 435
QY 122 LysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
Db 436 GGGCGAGAGGTCAACCTGACAAATTTCTGCTGTGTTCTTCTGGCAACAAATCGTCTC 495
QY 142 ThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
Db 496 AGGGTGGGACCGCGCAAGCTTGTGGACAGTGGCCCCAATGACCATATATTGTTTCTAC 555
QY 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
Db 556 AGTGACCATGGGGTCTCGCGTCTTGGAAATGCCTACGTATCATCTCTATGCTGAT 615
QY 182 AspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIle 201
Db 616 GACCTCGTAGATGCTCTGAAGAAGAGACATGCTGCGGGACCTTACAAAAGCCTGGTCTT 675
QY 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
Db 676 TACCTTGAAGCATGCCAATCTGGGACATCTTTGAGGGCTCTCTCGCGAATGACATAAAT 735
QY 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGlu 241
Db 736 GTGATGCGACCAACCGCTCAATGAGAGAGAGTAGTGGGGGACGTACTGCCCTGGC 795
QY 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
Db 796 GAGTTCGCCGAGCCCTCCACCGAGTAGTACACTTGTCTGGAGACCTGATAGTGTGCT 855
QY 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnIntyrHis 281
Db 856 TGGATGGAAGACAGTAGTATTTCCACAATCTGCGAAGTGAATCTCTCAAGCAGCAATACAAC 915
QY 282 ValValLysArgValGlySerAspValProGluThr-----SerHisVal 297

```

```

Db      916  TTGGTCAGAGTAGACACACG-----GTTCCAGATACATTACAGTATGCTCCCATGTC 969
Qy      298  CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
Db      970  ATGCAATATGGTTCATTGGAGTTGAATGTTAAGCATCTGTTTCTGATACATTGGCACAAAC 1029
Qy      318  ProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGly 337
Db      1030  CCTGCTAAGCATGACACACAGCTTACAGAACAACTCGTTGCCA--TCATTCTCAAG 1086
Qy      338  LeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaPro 357
Db      1087  GCTGTTATATCAGCGGACGCTGCTGCTTCTTCTGCGAGAGTACCGGAATTCGCCA 1146
Qy      358  MetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
Db      1147  GACAGCTACATGACAAAATGAGCTCGAGGAGTTCCTTGAAGTGATGGCCACAGG 1206
Qy      378  LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
Db      1207  TCTCATGTGTGACAGAGTGTGAGCTCATTTGGAAGCCCTCTCTTTGGCTCTGAGGACGGT 1266
Qy      398  LeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTyrAspCys 417
Db      1267  CCAAGGGTCTCGAAAGCCGTCGTCAGCTGCTGAGCCCTCTCTGTCGATGTTGGAGCTGT 1326
Qy      418  PheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeu 437
Db      1327  CTCAGTCCAGGTTCTGTTTTCGAGCGCAATGTCGGTGGTGGCGCAGTATGGGATG 1386
Qy      438  LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrVal 457
Db      1387  AAGCATGCGGTCCTCGCAACATCTGCAACGCTGCAACGCTGCAACGCTGCAACGCTGCG 1446
Qy      458  SerAlaIleGluGlnAlaCysSer 465
Db      1447  AAGGTCGTCGTCAGGCTTGCAAC 1470

```

RESULT 11

```

US-10-425-114-34046
; Sequence 34046, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 34046
; LENGTH: 1721
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLM01721A12_FLI
US-10-425-114-34046

```

Alignment Scores:

```

Pred. No.: 4,246-105 Length: 1721
Score: 1129.50 Matches: 225
Percent Similarity: 64.62% Conservative: 69
Best Local Similarity: 49.45% Mismatches: 146
Query Match: 46.03% Indels: 15
DB: 8 Gaps: 5

```

US-09-934-066-2 (1-466) x US-10-425-114-34046 (1-1721)

```

Qy      22  AlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLysSer 41
Db      143  CGCGAGGGTGAGTGGGAGCCGCTGATACGGATCGGACGGCGGAGGAGTGACGCTGCT 202
Qy      42  AlaLys-----GlyThrArgTrpAlaValLeuValAlaGly 53
Db      203  TCTGCTCCGGCTGCTGAGGATGATGAAGTGGGACGAGGTGGCCGCTGCTGCTGCCGA 262
Qy      54  SerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeu 73
Db      263  TCCTTCGGTATGGGAACCTACCGGCACCGCCGATGTGTCATGCGTACCATCACTACT 322
Qy      74  ArgLysGlyGluLeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPhe 93
Db      323  CAGAAGGAGAGTGAAGAAGAGAACATTTGGTGTTCATGATGATGATGATGATGATGAT 382
Qy      94  SerSerGluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyr 113
Db      383  AACATTCGAACCCCTAGGCGCTGGAGTTATCATTAACCATCTCTAAAGGAGCAATGTTAT 442
Qy      114  LysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnVal 133
Db      443  GATGGTGTTCGAAGGATTAACAGAGTGTACAGGTCTACTACTGAAAACTCTTTGCTGTC 502
Qy      134  LeuLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyPro 153
Db      503  CTCTTGGGCAATAGAAGTGCTACCACTCGAGGGAGTGAAGAAGTGATAGACAGCAACCT 562
Qy      154  AsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetPro 173
Db      563  AATGACCATCTCTCATCTATTACTCATGATCGGGGTCCTGGAGTCTTCTGATGATGCGG 622
Qy      174  ThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArg 193
Db      623  AACCTGCCCATATCTGATGTGTCGTCATCTCAAGGCTTGAAGAAGAGCAAGTCTTTCAG 682
Qy      194  LysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGlu 213
Db      683  AATAGCTACTCAAAATGTTATATATGTTGAAGCATGTGAAGTGGCAGTATCTTTGAG 742
Qy      214  GlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSer 233
Db      743  GGGTTAATGCCTGAAGATCTAAATATTTATGTCACACAGCCTCAAACTTGAAGT 802
Qy      234  SerTrpGlyValTyrCysProGluSerTyrProProProProSerGluIleGlyThrCys 253
Db      803  AGTTGGGAACCTTACTGCTGGGATGGAACCATCCCTCTCTGAGTACATTTACCTGT 862
Qy      254  LeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLys 273
Db      863  TTAGGTGATTTGTACAGTGTCTTGGATGGGAAGACAGTGAACCCCACTTAAGGAAG 922
Qy      274  GluThrLeuGluGlnGlnTyrHisValValLysArgValGlySerAsp-----Val 291
Db      923  GAAACCATCAAGCATCAGTATGAAGTGGTGAACCAAGCAAGCTCAAACTCAATTAAGTAC 982
Qy      292  ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer 311
Db      983  AAAGAGGTTCTCATGTTATGGAGTATGTTGATAGACCTTCAAGGATGAGAAGCTTTC 1042
Qy      312  SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSer 331
Db      1043  TTTTATCAAGGTTTTCATCTCTCTAATGCCAATCTGCAACATCTGCAACATCTGCTGTC 1102
Qy      332  ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArg 351
Db      1103  CCA-----AAGGGTGCATCTCAATCAAGAGATGCTGATCTCTCTTCTTCACTGTGAAG 1153
Qy      352  LysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 371
Db      1154  AGGTACGAGCAGCTTAAATGGGGGACCGAAGAGAGCTGAGGGCTCTCATAGAGATTAAA 1213
Qy      372  AspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSer 391

```

```

Db 1214 GAAACTGTTGACGACAGGAGCATCTTGACAGCAGTATCGATTTCGTGGAGGCTCGTC 1273
Qy 392 ValLysGlnThrAsnValLeuAsnLeuThrSerThrArgThrThrGlyGlnProLeu 411
Db 1274 TTCGGATTGAGAAAGGCGCTTCATGCTGTCAGGCTTTAGAACCTCTGGTCAACCATTA 1333
Qy 412 ValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAla 431
Db 1334 GTCGATCAGTGGGACTCTTTGAAGAGGAGTGGTGGGATTTTCGAGTCCGAGTGGGTCA 1393
Qy 432 ThrValHisThrGlyLeuLysThrGlyAlaLeuAlaAsnLeuCysAsnMetGlyVal 451
Db 1394 CTTACTCAGTACGGCATGAGTACATGAGAGCGCTTTCGAACATCTGCAACAGTGGTATA 1453
Qy 452 ---AspValLysGlnThrValSerAlaIleGluAlaCysSer 465
Db 1454 TCTGAGATGAATGAGGGAGTCGAGCATC---AGCGCTTGCAGC 1495

RESULT 12
US-10-425-114-27715
: Sequence 27715, Application US/10425114
: GENERAL INFORMATION:
: APPLICANT: Liu, Jingdong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovacic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: CURRENT APPLICATION NUMBER: US/10/425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128
: SEQ ID NO 27715
: LENGTH: 1809
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB4739-001-A9_FLI
US-10-425-114-27715

Alignment Scores:
Pred. No.: 1.17e-104 Length: 1809
Score: 1125.50 Matches: 225
Percent Similarity: 64.40% Conservative: 68
Best Local Similarity: 49.45% Mismatches: 147
Query Match: 45.86% Indels: 15
Db: 8 Gaps: 5

US-09-934-066-2 (1-466) x US-10-425-114-27715 (1-1809)
Qy 22 AlaGluSerArgLysThrGlnLeuLeuAsnAspAsnValGluSerSerAspLysSer 41
Db 161 GCGGAGGAGTGGGAGCGCTGATACGAGTCCGACGCGGAGGAGGAGTGGCTGCT 220
Qy 42 AlaLys-----GlyThrArgTrpAlaValLeuValAlaGly 53
Db 221 TCTGCTCGGCTGCTGAGGATGATGAAGTGGGACGAGGTGGCGGCTGCTGCTGCGCGA 280
Qy 54 SerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspLysCysHisAlaTyrGlnIleLeu 73
Db 281 TCCTTTCGTTATGGGAACCTACCGGCACCGAGCGGATGTCGCCATGCTACCAGATCTG 340
Qy 74 ArgLysGlyGlyLeuLysAspGluAsnIleValPheMetThrAspAspIleAlaPhe 93
Db 341 CAGAAGGGAGGAGTGAAGAGGAGAACATTTGCTGTCATGATGATGATGATGATGATGAT 400
Qy 94 SerSerGluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyr 113
Db 401 AACATCTCGAACCTAGCGCTGGAGTTATCATTAACCATCTCTAAAGGAGCAATGTTAT 460

```

```

114 LysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnVal 133
Db 461 GATGGTGTCCAAAGGATTACACAGGTGATCAGTCACTACTGAAACCTCTTTTGGTGTGTC 520
Qy 134 LeuLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValLysSerGlyPro 153
Db 521 CTCTTGGGCAATAGAAAGTGTACCACTGAGGAGTAAAGAGGTGATAGACAGCAACCT 580
Qy 154 AsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetPro 173
Db 581 AATGACCACATCTCTATCTACTCAGATCATGGGGTCTCTGGAGTCTTGGTATGCCG 640
Qy 174 ThrGlyAspGluValMetalLysAspPheAsnGluValLeuGluLysMetHisLysArg 193
Db 641 AACCTGCCATATCTGATGCTGCTGCTCATCAAGGTCTTGAAGAGTCTTGAAGAGAGCATGCTTCC 700
Qy 194 LysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGlu 213
Db 701 AATAGCTACTCAAAATGTTATATATGTTGAAGCATGTGAAAGTGGCAGTATCTTTGAG 760
Qy 214 GlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSer 233
Db 761 GGGTAAATGCTCAAGATCTAATATATTATGTCAACACAGCTCAAACTCAGTTGAAAT 820
Qy 234 SerTrpGlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCys 253
Db 821 AGTTGGGAACTTACTGCCCTGGATGGAACCATCACCCCTCTCTGAGTACATTTACCTGT 880
Qy 254 LeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLys 273
Db 881 TTAGTGATTTGTCAGATGTTTCTTGATGGAAGACAGTGAACCCCACTCTAAAGAG 940
Qy 274 GluThrLeuGlnGlnTyrHisValValLysArgValGlySerAsp-----Val 291
Db 941 GAAACCATCAAGCATCAGTATGAAGTGTGAACACAGACCTCAAACTCAATAAGTAC 1000
Qy 292 ProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSer 311
Db 1001 AAAGAGGTTCTCATGTTATGAGTATGTTGATGAAGACCTTCAAGGATGAGAGCTTTC 1060
Qy 312 SerTyrIleGlyArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerSer 331
Db 1061 TTTTATCAAGGTTTGTATCTGCTTAATGCCAATGTCGAACATGCTGGTGTGGCCAGGC 1120
Qy 332 ProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLysTyrLeuGlnArg 351
Db 1121 CCA-----AAGGGTGCAGTCAATCAAGAGATGCTGATCTTCTCTCATGTGGAAG 1171
Qy 352 LysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeu 371
Db 1172 GGGTACGAGCAGTTAAATGGGGGACCGAAGAGCTGAGGCTCTCATAGAGATTTAA 1231
Qy 372 AspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSer 391
Db 1232 GAAACTGTTCCAGCACAGGAAGCATCTTGACAGCAGTATCATGTTCTGGTGGAGGCTCGTC 1291
Qy 392 ValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeu 411
Db 1292 TTCGGATTGAGAAAGGCGCTTCGATGCTTGAAGCTGTTAGAACCCTCTGGTCAACCATTA 1351
Qy 412 ValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAla 431
Db 1352 GTCGATCAGTGGGACTCTTTGAAGAGGAGTGGTGGGATTTTCGAGTCCGAGTGGTGTCA 1411
Qy 432 ThrValHisThrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyVal 451
Db 1412 CTTACTCAGTACGGCATGAAGTACATGAGAGCGCTTTCGAACATCTCTCAACAGTGGTATA 1471
Qy 452 ---AspValLysGlnThrValSerAlaIleGluAlaCysSer 465
Db 1472 TCTGAGATGAATGAGGGAGTCGAGCATC---AGCGCTTGCAGC 1513

```

RESULT 13

US-10-425-114-35501

; Sequence 35501, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35501
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROB73045E08_FLI
US-10-425-114-35501

Alignment Scores:
Pred. No.: 5,97e-100 Length: 1527
Score: 1078.50 Matches: 207
Percent Similarity: 66.75% Conservative: 64
Best Local Similarity: 50.99% Mismatches: 128
Query Match: 43.95% Indels: 7
DB: 8 Gaps: 3

US-09-934-066-2 (1-466) x US-10-425-114-35501 (1-1527)

Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
Db 2 GCGACATCTGCATGATACCATGATGAGAGGCGGACATCAAGCATGAGAACATC 61
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerGluAsnProArgProGlyValIle 103
Db 62 ATTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121
Qy 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGlu 123
Db 122 ATCAACCATCCCGAGGTGGCGACGCTATGCTGGGTGCCAAGGATTACACTGGTCTGA 181
Qy 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143
Db 182 GATGTCAACGTCGACAAATTTCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
Qy 144 GlyAsnGlyValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 242 GCGAGCGCGAGGTGTGACAGTGGCCCAATGATCATATCTTTGTTCTACAGTGC 301
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 302 CATGGGGTCTCTGGGTCTCTGGAATGCTACGATATCCATATCTATGTTGATGACCTC 361
Qy 184 AsnGluValLeuGluLysMethisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 362 GTAGATGTCTTGAAGAAGATGATGCTCCCGGACCTACAAAGCCCTGTTTACCTT 421
Qy 204 GluAlaCysGluSerGlySerMetPheGluClyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 422 GAAGCATGCGAATCTGGGAGCATCTTTGAGGCGCTCTCCGCGAATGACATCAATGTAT 481
Qy 224 AlaValThrAlaAlaAsnSerLysLysSerTrpGlyValTyrCysProGluSerTyr 243
Db 482 GCGACCCCGCTCAATATGACAGCAGCAGTGGGGAGCTATGCTGCTGCTGCTGCTGCT 541
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 542 CCGAGCCCTCCACCGAGTATGACACTTGTGTTGGGAGACCTGTACAGTGTGCTTGGATG 601

Qy 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHisValVal 283
Db 602 GAAGACAGTGAATTCACAACTCTGCAATGTAATCCTCAAGCAGCAGTACAAAGTGTGTC 661
Qy 284 LysArgArgValGlySerAspValProGluThr-----SerHisValCysArg 299
Db 662 AAGGATAGGACAGCG-----GTTTCATGACACGTTTCAGCTATGGTTCCTCATGTATGCA 715
Qy 300 PheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu 319
Db 716 TATGGCGCATGAGAGCTGAATGTTTCAGCGTTTGTTCGTACATGTCACAGACCTGCT 775
Qy 320 AsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsnSerGlyLeuVal 339
Db 776 AAGCATGGCAACACTTTTATAGAAGTAACCTCACTGCCA---TCGTTCTCAAGAGCCGTT 832
Qy 340 AsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGly 359
Db 833 ATACAGCGCAGCGCTGACCTTGTCTACTTCTGCAAGATACCGCAATATGCTGTGATAGC 892
Qy 360 SerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGln 379
Db 893 TCGCATGCGAAAAATGAGGCTCGGAAGAACTGCTTGAAGTGTATGGCCACAGGTCTCAT 952
Qy 380 IleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsn 399
Db 953 GTTGACAGCAGCGCTGAGCTCATTTGGAAGCCTTCTTTGGCTCGAGGAGCGGTCCAAG 1012
Qy 400 LeuLeuThrSerThrArgThrThrGlyGlnProLeuValAlaAspTrpAspCysPheLys 419
Db 1013 GTTCTGAAAGCGCTCGCTGCACCTGCTGAACCTCTGTTGATGATGAGGCTGCTCAAG 1072
Qy 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 1073 TCCATAGTTTCGCACTTTTCAGGCGCGATCGGGTCTCTTGGCGCAGTATGGATGAAGCAC 1132
Qy 440 ThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAla 459
Db 1133 ATCGATGCTTCGGAACATGTGCAAGCTGCGCTCTCTCTGGAAGCAGTGTGCAAGGTG 1192
Qy 460 IleGluGlnAlaCysSer 465
Db 1193 GCGGCTCAGCGCTGCGACG 1210

RESULT 14

US-10-424-599-49438 Application US/10424599
; Sequence 49438, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 49438
; LENGTH: 2087
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_1564C.1
US-10-424-599-49438

Alignment Scores:
Pred. No.: 1.32e-99 Length: 2087
Score: 1077.00 Matches: 210
Percent Similarity: 64.83% Conservative: 72
Best Local Similarity: 48.28% Mismatches: 133
Query Match: 43.89% Indels: 20
DB: 8 Gaps: 4

US-09-934-066-2 (1-466) x US-10-424-599-49438 (1-2087)

Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 379 CACTGATCGGATGAAGTGGGAACACAGATGGCGGTCTCTCGTGGTTCACACGGCTAC 438

Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77
Db 439 GGAACATACAGGATCAAGCAGATGTGTCATCGCTACCATGCTGCTGATAAAAGTGA 498

Qy 78 LeuLysAspGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn 97
Db 499 CTAAAGAGAGACATAGTGGTGTATGTAGCATGACATAGTACCAACGAGTGGAT 558

Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysValPro 117
Db 559 CCTAGACATGGATCATCATCAACACCTCAGGAGAGATCTGTATGCTGTCTTCT 618

Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsn 137
Db 619 AAGGATTACACCGGTGATAATGTGACGACGAGAACCTCTTGTCTGTATTCTTGGAGAC 678

Qy 138 GluSerGlyValThrGlyLysGlyValValLysSerGlyProAsnAspAsnIle 157
Db 679 AAGAGTAATTTGAAGGAGGAAGTGGCAAGTCAACAGCAAAACCCGAGGACAGATA 738

Qy 158 PheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 739 TTTATATACATCTGATCATGAGGCTCTGGAATACTTGGGATGCCAAACATGCCATAC 798

Qy 178 ValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLysTyrAsn 197
Db 799 CTTTATCCCATGGATTTATGTATGTCTTGAAGAAGAAACATGCCATCTGGAAGTTACAAG 858

Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys 217
Db 859 GAGATGTTATATACGTGGAGCTTGTGAAGTGGGAGCGTGTGTGAGGTATTAATGCCT 918

Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 919 AAGGATCTGAATATTTATGTCACAACTGCATCAAAATCCACAGAGAATAGTTGGGAACT 978

Qy 238 TyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspThr 257
Db 979 TATGTCTCTGGAATGGATCTCTTCCACCTCCAGAGTACATCATCTCCCTAGGGATTG 1038

Qy 258 PheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGlu 277
Db 1039 TACAGCGTGTGGATGGAGATAGTAGGCTCACAAATCTAAAGGGAATCCGTGAAA 1098

Qy 278 GlnGlnTyrHisValValLysArgArgValGly-----SerAspValProGluThrSer 295
Db 1099 CAACAATACAAATCGGTAAAGAACCGACATCAAAATTTCAACAACTATGCGATGGGTCT 1158

Qy 296 HisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGly 315
Db 1159 CATGTGTGATGATATGTTGATCAACACATCACAGCTGAAAGCTTTATTTATCAACAAAGT 1218

Qy 316 ArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsn 335
Db 1219 TTTGATCCTGCCACTGTGAACCTC-----CTCCACAAACAA 1254

Qy 336 SerGlyLeu-----ValAsnProArgAspIleProLeuLeuTyrLeu 349
Db 1255 GGCAGGCTAGAACTAAATGGAAGTGTAAACCAAGAGATGACAGAACTTTTGTTCATG 1314

Qy 350 GlnArgLysIleGlnLysAlaPrometGlySerLeuGluSerLysGluAlaGlnLysLys 369
Db 1315 TGGCAATGTATCAGATCAAAACATCATCAGTGAAGAAATTAAGACAGACATCTCCCAACAA 1374

Qy 370 LeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArg 389
Db 370 LeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArg 389

Db 1375 ATTGGCGGAGACACTGAAGCATAGGAACACATAGATGGTAGCGTGAATTCATTGGAGTT 1434

Qy 390 LeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGln 409
Db 1435 TTACTGTATGTGACCAAGAAAGTTCTTCTGTTCACAAATCCGTGAGGGCTCTCGTTGC 1494

Qy 410 ProLeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCys 429
Db 1495 TCCCTTTTGTGATGACTGGACATGCCATAAATCAATGTTCCGGTGTGTTGAACACTCACTGT 1554

Qy 430 GlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMet 449
Db 1555 GGACACTGACTCAGTATGTCATGAACACATCGGACATTCGCCAACATTTGCAACAGT 1614

Qy 450 GlyValAspValLysGlnThrValSerAlaIleGluAlaCys 464
Db 1615 GCGCTT-----TCTGAGGCTCCATGGAAGAGCTTGT 1647

RESULT 15

US-10-425-114-10538
; Sequence 10538, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 10538
; LENGTH: 1651
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700942962_FL1
US-10-425-114-10538

Alignment Scores:

Pred. No.:	2.58e-98	Length:	1651
Score:	1063.00	Matches:	200
Percent Similarity:	68.24%	Conservative:	60
Best Local Similarity:	52.49%	Mismatches:	119
Query Match:	43.32%	Indels:	2
DB:	8	Gaps:	2

US-09-934-066-2 (1-466) x US-10-425-114-10538 (1-1651)

Qy 85 ValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIleIle 104
Db 1 GTTTTATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60

Qy 105 AsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThrLysGluAla 124
Db 61 AACAAACACATGGAGGTGATGTTTATAAAGGATTTCCAAAGGATTAACCCGCGAAGAT 120

Qy 125 ValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGly 144
Db 121 GTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

Qy 145 AsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHis 164
Db 181 AGTGGGAGAGTTGTGGACAGTGTGTCCTGATGATGATGATGATGATGATGATGATGATGAT 240

Qy 165 GlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsn 184
Db 241 GGAGGCTCTGGGTGCTCGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

Qy 185 GluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrValGlu 204

Db	301	GAAGTCTTGGAGAAAACCATGCTTCTGGAAACATATAAAACCTAGTATTTTATCTGGAG	360
Qy	205	AlaCysGluSerGlySerMetPheGluGlyIleLeuLysAsnLeuAsnIleTyrAla	224
Db	361	GCATGTGAATCTGGGAGTATCTTTGAAGGTCTTCTTCCCTGAAGATATCAATATTTATGCA	420
Qy	225	ValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyrPro	244
Db	421	ACCACCTGCTCCATTCGAGAGAAAGTAGTTGGGGAACATATTTGCCCGGGGAGTATCCT	480
Qy	245	ProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGlu	264
Db	481	AGTCCTCCCCAGAAATATACAACCTGTTTGGGTGACTTGTACAGTGTGTCTGTGGATGCAA	540
Qy	265	AspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHisValValLys	284
Db	541	GACAGTGCACACACAATTTGCGAACAAGAACTCTCCACCAACAATATATAAATGGTTAAA	600
Qy	285	ArgArg---ValGlySerAspValProGluThrSerHisValCysArgPheGlyThrGlu	303
Db	601	GAGAGGACTATATCTGGAGATTCTACTATGTGGCTCTCACGTGATGCAGTATGGTGATCTA	660
Qy	304	LysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGluAsnAspAsnPhe	323
Db	661	GGGCTTAGCAGAGATGTTCTCTCCATATTTGGGTACAGATCCTGCTAAATGATAATTC	720
Qy	324	ThrPheThrGluSerPheSerSerProIleSerAsnSerGlyLeuValAsnProArgAsp	343
Db	721	ACTTTGTGGATGAA---AACTCCTTATGTCTACCTTCAAAACCAAGTCACCAACGTCAT	777
Qy	344	IleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGlySerLeuGluSer	363
Db	778	GCTGATCTCATCCATTTTGGGATAGTTCCTCCGCAAGCTCTCAGGGTCTCTCAGGAA	837
Qy	364	LysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSer	383
Db	838	ANTACAGCTAGAAACAAGATTTTGGAGCAATCTCTACAGAATGCATGTAGACAACAGT	897
Qy	384	IleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSer	403
Db	898	GTAAACATCATTTGGAGACTTTATTGTCATTTGAAAGGGTCCAGAACTACTCAACGCT	957
Qy	404	ThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLysThrLeuValAsn	423
Db	958	GTTACCGCGGTGATCGGCATTTGTGTATGACTGGCACTGCCTGAAACCACTGGTGAGG	1017
Qy	424	SerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyrThrGlyAlaLeu	443
Db	1018	ACTTTTGTGAGACACATTTGGATCTCTGCTCAATACGGGATGAACACATGAGTGCCTTT	1077
Qy	444	AlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAlaIleGluAla	463
Db	1078	GCAAACATCTCAATGTAGGATATAAGAATGAACAAATGGCTGAGGCTTCAGCAACAAGCT	1137
Qy	464	Cys 464	
Db	1138	TGT 1140	

Search completed: May 29, 2003, 20:40:32
Job time : 719 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: May 29, 2003, 18:04:19 ; Search time 1602 Seconds
(without alignments)
4711.047 Million cell updates/sec

Title: US-09-934-066-2
Perfect score: 2454
Sequence: 1 MSSPLGHFQILVFLHALLIF.....CNMGVDVKQTVSAIEQAQCSM 456

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US0934066/runat_19052003_163700_24361/app_query.fasta_1.647
-DB=EST -QPMF=fastap -SUFFIX=p2n.est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0934066.ecgn_1.1.2874 -runat_19052003_163700_24361 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estio:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1204.5	49.1	1981	11	AV104034	AV104034 Zea mays
2	1089.5	44.4	1631	11	AV110063	AV110063 Zea mays
3	968	39.4	541	10	AV559545	AV559545 AV559545
c 4	924	37.7	606	10	AV557684	AV557684 AV557684
5	922	37.6	760	14	BQ871979	BQ871979 OG113F11
6	921	37.5	746	14	BQ871135	BQ871135 OG110P16
7	907	37.0	796	12	BG645154	BG645154 EST506773
8	904	36.8	711	12	BG646037	BG646037 EST507656
9	903	36.8	696	12	BG441524	BG441524 GA__ca001
c 10	896	36.5	766	12	BG123960	BG123960 EST469606
c 11	893	36.4	585	10	AV559121	AV559121 AV559121
12	869	35.4	776	12	BG581593	BG581593 EST483328
c 13	867	35.3	561	10	AV559478	AV559478 AV559478
14	855	34.8	538	10	AV555212	AV555212 AV555212
15	848	34.6	535	10	AV562667	AV562667 AV562667
16	818	33.3	669	10	AV940380	AV940380 AV940380
17	817	33.3	574	10	AV831492	AV831492 AV831492
18	815.5	33.2	801	13	BI308524	BI308524 EST529934
19	814	33.2	655	14	BQ871920	BQ871920 OG113C24
20	804	32.8	733	14	BU028909	BU028909 QH6823.Y
21	802	32.7	679	10	AV926752	AV926752 AV926752
c 22	784	31.9	506	10	AV557191	AV557191 AV557191
23	782	31.9	719	14	BQ841709	BQ841709 WHE2984.C
24	758	30.9	431	10	AV564659	AV564659 AV564659
25	752	30.6	617	14	BQ579190	BQ579190 WHE2966.F
26	752	30.6	718	12	BG129376	BG129376 EST475022
27	749	30.5	845	12	BF276350	BF276350 GA__Eb002
c 28	745	30.4	836	12	BF263324	BF263324 HV_Cea000
29	743	30.3	535	10	AV561906	AV561906 AV561906
c 30	742	30.2	646	14	BU025783	BU025783 QHG11G16
31	739	30.1	638	10	BE660642	BE660642 1043 Gmax
32	738	30.1	654	10	AW224404	AW224404 EST298291
33	732.5	29.8	1345	11	AW106221	AW106221 Zea mays
34	730.5	29.8	716	10	BE660892	BE660892 246 Gmax
35	730	29.7	919	14	BQ881734	BQ881734 AGENCOURT
36	725	29.5	768	14	BU004552	BU004552 OGG5111.Y
37	722.5	29.4	1057	9	AL544478	AL544478 AL544478
38	716.5	29.2	681	10	AW224405	AW224405 EST298292
39	716	29.2	661	14	BQ148226	BQ148226 NF063A11F
40	714.5	29.1	765	12	BG591158	BG591158 EST499000
41	704	28.7	595	9	AJ436163	AJ436163 AJ436163
42	700.5	28.5	861	14	BQ936258	BQ936258 AGENCOURT
43	696.5	28.4	755	14	BU002036	BU002036 OGG30A09
44	696	28.4	865	13	BI953500	BI953500 HVSMEN001
45	695	28.3	600	14	BQ997181	BQ997181 OGG14P12

ALIGNMENTS

RESULT 1
AV104034
LOCUS
DEFINITION Zea mays PC0131778 mRNA sequence.
ACCESSION AV104034
VERSION AV104034.1 GI:21207112
KEYWORDS HPC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1981)
REFERENCE
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 1981)
 AUTHORS Coe, E.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 FEATURES Location/Qualifiers
 source 1..1981
 /organism="Zea mays"
 /db_xref="taxon:637688"
 /clone_lib="PCO131778"
 /clone="PCO131778"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 480 a 509 c 505 g 487 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4,28e-137 Length: 1981
 Score: 1204.50 Matches: 227
 Percent Similarity: 69.16% Conservative: 69
 Best Local Similarity: 53.04% Mismatches: 125
 Query Match: 49.08% Indels: 7
 DB: 11 Gaps: 3

US-09-934-066-2 (1-466) x AY104034 (1-1981)

Qy 42 AlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyrArg 61
 Db 203 GCGTCGGAGCCGGTGGCGGTCATCGCGGCTCCAGCGCTACTACAACTACCGC 262
 Qy 62 HisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGlu 81
 Db 263 CACGAGCGGACATCGCCATGATACACAGATCATGAAGAGCGGCGCTTAAGACGAG 322
 Qy 82 AsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGly 101
 Db 323 AACATCGTGTCTATGATGACATGACATCGCGCATAGCCCGGAAATCCGAGGCGCTGT 382
 Qy 102 ValIleLeuAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrThr 121
 Db 383 GTCATCAATCAATCCCGAGGCTGGCGACGCTCTATGCTGGGGTGCCCAAGGATTACACT 442
 Qy 122 LysGluAlaValAsnValGlnAsnPhenTyrAsnValLeuLeuGlyAsnGluSerGlyVal 141
 Db 443 GGGCGAGAGGTCAACGTCGACATTTCTTCGCTGTCTGTTGGCAACAAACTGCTCTC 502
 Qy 142 ThrGlyGlyAsnGlyLysValLysSerGlyProAsnAspAsnIlePheIleTyrTyr 161
 Db 503 AGGGGTGGGAGCGCAAGCTTGTGGACAGCTGCCCAATGACCATATATTGTTTCTAC 562
 Qy 162 AlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLys 181
 Db 563 AGTGACCATGGGGTCTCGGCTCTTGAATGCCATACGTATCATATCTATGCTGAT 622
 Qy 182 AspPheAsnGluValLeuLysMetHisLysArgLysLysTyrAsnLysMetValIle 201
 Db 623 GACCTCGTATGATCTCTGAAGAAGAGCATGCTCGCGGACCTACAAAGCGCTGCTTT 682
 Qy 202 TyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsn 221
 Db 683 TATCTTGAAGCATGCCAAATCTGGAGCATCTTTGAGGCGCTCTCCCGAATGACATAAT 742
 Qy 222 IleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGly 241

Db 743 GTGTATCGGACCACCGCGTCAATGTCAGAGGAGTAGCTGGGGAGCTACTGCCCTGCC 802
 Qy 242 SerTyrProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSer 261
 Db 803 GAGTTCGCGAGCCCTCCCGGAGTATGACACTTGTGGAGACCTGTATAGTGTGCT 862
 Qy 262 TrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnGlnTyrHis 281
 Db 863 TGGATGGAAGACAGATGATTTCCAAATCTCGCAACTGAATCTCTCAAGCAGCAATACAA 922
 Qy 282 ValValLysArgArgValGlySerAspValProGluThr-----SerHisVal 297
 Db 923 TTGGTCAAGATAGACAGCG-----GTTACAGATACATTCAGCTATGCTCCCATGTG 976
 Qy 298 CysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsn 317
 Db 977 ATCAATATGTTCAATTTGGAGTTGAATGTTAAGCATCTGTTTCGPACATTTGCCACAA 1036
 Qy 318 ProGluAsnAspAsnPheThrGluSerPheSerSerProIleSerAsnSerGly 337
 Db 1037 CCTGCTAACCATGACAAACAGCTTTATAGAAGACAACTCGTTGCCA---TCGTTCTCAA 1093
 Qy 338 LeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaPro 357
 Db 1094 GCTGTTAATCAGCGCGAGCGTGTCTACTTCTGGCAGAGTACCGGAATTTGCA 1153
 Qy 358 MetCysSerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArg 377
 Db 1154 GACAGTCACCTGAGAAAATGAAGCTCGAAGAGTGTCTGAGTGTAGTGGCCACAGG 1213
 Qy 378 LysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnVal 397
 Db 1214 TCTCATGTTTCACAGCAGCTGTGTGAGCTCATTTGGAAGCCTTCTCTTGGCTCTGAGGCGGT 1273
 Qy 398 LeuAsnLeuLeuThrSerThrArgThrGlnProLeuValAspAspTrpAspCys 417
 Db 1274 CCAAGGTTCTGAAAGCCGTCGTCGAGCTGTGTGAGCCTCTGTGTCGATGATTGGAGCTGT 1333
 Qy 418 PheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeu 437
 Db 1334 CTCAAGTCCACGCTGCTGCTACTTTGAGCGGCAATGTGGTCTGTTGGCGCAGTATGGGATG 1393
 Qy 438 LysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrVal 457
 Db 1394 AAGCACATGCGGTCTTCGCAACATCTCGAAGCTGGCATCTCTCTCAAGCAGTGTGCTG 1453
 Qy 458 SerAlaIleGluGlnAlaCysSer 465
 Db 1454 AAGTCTGCTCTCAGGCTTGACCC 1477

RESULT 2
 AY110063
 LOCUS
 DEFINITION Zea mays CL863_1 mRNA 'sequence'.
 ACCESSION AY110063
 VERSION AY110063.1 GI:21214138
 KEYWORDS HTG.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 Clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 1651)
 AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 1651)
 AUTHORS Coe, E.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA

/clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 177 a 88 c 128 g 148 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9,99e-109 Length: 541
 Score: 968.00 Matches: 180
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.45% Indels: 0
 DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x AV559545 (1-541)

QY 96 GluAspProGluValIleLeuAsnLysProAspGlyGluAspValTyLysGly 115
 Db 2 GAGATCCTAGCCCTGGAGTTATCATTAATTAACCGATGGAGAGATGTTTATAAGGA 61
 QY 116 ValProLysAspTyThrLysGluAlaValAsnValGlnAsnPhetYrAsnValLeu 135
 Db 62 GTTCCTTAGGACTACACTAAGAGCTGTTAATGTCAAACTTCTACAATGTGTACTT 121
 QY 136 GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAsp 155
 Db 122 GGAATGAAGTGGCGTCAAGAGGAAATGCAAAAGTGTGAAAGTGGTCTTAATGAT 181
 QY 156 AsnIlePheIleTyThrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
 Db 182 ATATCTTCATATATGCTGACCATGGAGCTCTGCTTAAATAGCGATGCCACTGGT 241
 QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGlyLysMetHisLysArgLysLys 195
 Db 242 GATGAAGTTATGCAAAAGATTTCATGAAGTCTTGGAGAGATGTCATTAAGAGAAAA 301
 QY 196 TyrAsnLysMetValIleTyThrValGluAlaCysGluSerGlySerMetPheGluGlyLe 215
 Db 302 TACAACAAGATGGTGATCTATCTGAAGCATGTGAATCAGGAAGTATGTTTGAAGGGAT 361
 QY 216 LeuLysLysAsnLeuAsnIleTyThrAlaValThrAlaAlaAsnSerLysGluSerSerT 235
 Db 362 TTAAGAAAAATCTCAACATATACGAGTGTGCTGCTCTCTCTCTCTCTCTCTCTCT 421
 QY 236 GlyValTyCysProGluSerTyProProProSerGluIleGlyThrCysLeuGly 255
 Db 422 GGAGTTTACTGTCTGAGTCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 481
 QY 256 AspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThr 275
 Db 482 GATACATTAGCATCTCTTGGCTTCAGGCACAGTGCACCTTCATGACATGACGAGAGACT 541

RESULT 4
 AV557684/C
 LOCUS
 DEFINITION
 AV557684 Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone SQ075c12f 3', mRNA sequence.
 AV557684
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 606)
 Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)
 20363093

COMMENT

Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

Location/Qualifiers
 1..606
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ075c12f"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 158 a 110 c 124 g 214 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.18e-103 Length: 606
 Score: 924.00 Matches: 177
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 37.65% Indels: 0
 DB: 10 Gaps: 0

US-09-934-066-2 (1-466) x AV557684 (1-606)

QY 290 AspValProGluThrSerHisValCysArgPheGlyThrGluLysMetLeuLysAspTyr 309
 Db 606 GATGTACACAGAGACTTCTCATGTATGCCGTTTCGGAACAGAGAAGATGCTTAAGATTAT 547
 QY 310 LeuSerSerTyRileGlyArgAsnProGluAsnAspAsnPhetThrPheThrGluSerPhe 329
 Db 546 CTTTCTCTTACATTGGAAGAAATCTGAAACAGATAACTTCACTTTCACGGAATCCTTT 487
 QY 330 SerSerProIleSerAsnSerGlyLeuValAsnProArgAspIleProLeuLeuTyRLeu 349
 Db 486 TCCTCACCAATCTCTAATTCCTGGCTTGGTCAATCCGCGGATATTCCTCTGTATACCTC 427
 QY 350 GlnArgLysIleGlnLysAlaProMetGlySerLeuGluSerLysGluAlaGlnLysLys 369
 Db 426 CAGAGAAGATTCAAAAGCTCCAATGGCATCCTTGAAGCAAGAGAGCTCAGAAGAAA 367
 QY 370 LeuLeuAspGluLysAsnHisArgLysGlnIleAspGlnSerIleThrAspIleLeuArg 389
 Db 366 TTGCTTGACGAAAAGAAATCATAGGAACAAATCGATCAGAGCATTTACAGACATTCCTCGG 307
 QY 390 LeuSerValLysGlnThrAsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGln 409
 Db 306 CTTTCAGTTAAACAAACCAATGCTTAAATCTCTTAACCTCCACAGAACACACAGACAG 247
 QY 410 ProLeuValAspAspTrpAspCysPheLysThrLeuValAsnSerPheLysAsnHisCys 429
 Db 246 CCTCTGTACAGCATGGGATTCCTCAAGACTCTAGTTAATAGCTTCAAGAAATCAGTGC 187
 QY 430 GlyAlaThrValHisTyRileGlyLeuLysTyThrGlyAlaLeuAlaAsnIleCysAsnMet 449
 Db 186 GGTGCAACGGTGCATTTACGGATTGAAGTATACAGAGCGCTTGCACATATCTGCCAATATG 127
 QY 450 GlyValAspValLysGlnThrValSerAlaIleGluGlnAlaCysSerMet 466
 Db 126 GGAGTGGATGTGAAGCAAACTCTTTACGCCATTTGAACAGCTTGTTCGATG 76

RESULT 5
 BQ871979
 LOCUS
 DEFINITION
 BQ871979 lettuce salinas
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE

BQ871979 760 bp mRNA linear EST 15-AUG-2002
 OG113F11.yg.abl OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 OG113F11, mRNA sequence.
 BQ871979
 BQ871979.1 GI:22258529
 EST.
 Lactuca sativa.

```

ORGANISM      Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.

REFERENCE
AUTHORS      1 (bases 1 to 760)
              Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
              Lin,H., van Damme,M., Lavelle,D., Chevallier,P., Ziegler,J., Ellison
              ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
              Church,S., Jackson,L. and Bradford,K.
              Lettuce and Sunflower ESTs from the Compositae Genome Project
              http://compgenomics.ucdavis.edu/
              Unpublished (2002)
              Contact: Alexander Kozik [R.W.Michelmore]
              Department of Vegetable Crops, R.W.Michelmore Lab
              University of California at Davis (UCD)
              Asmudson Hall, UCD, Davis, CA 95616, USA
              Tel: 1-(530)-742-1742
              Fax: 1-(530)-752-9659
              Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
              belongs to contig QG_CA_Contig7059, see http://cgpdb.ucdavis.edu/
              for details.
              Plate: QG110 row: F column: 11.

FEATURES
source       Location/Qualifiers
              1..760
                /organism="Lactuca sativa"
                /cultivar="Salinas"
                /db_xref="taxon:4236"
                /clone="QG113f11"
                /clone_lib="QG_ARCDI lettuce salinas"
                /lab_host="E.coli"
                /note="Vector: pBRCNDSfiAB; The library was constructed
                from 10 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_LIB-QG_ARCDI lettuce salinas
                TAG_TISSUE=roots
                TAG_SEQ=GTTCACGGG"

BASE COUNT   232 a 131 c 190 g 207 t
ORIGIN
Alignment Scores:
Pred. No.:    7.8e-103      Length: 760
Score:        922.00       Matches: 162
Percent Similarity: 78.86%   Conservative: 32
Best Local Similarity: 65.85% Mismatches: 52
Query Match:  37.57%       Indels: 0
DB:           14          Gaps: 0

US-09-934-066-2 (1-466) x B0871979 (1-760)

QY 44 GlyThrargTrpAlaValLeuAlaGlySerAsnGluTyrTrpAsnTyrArgHisGln 63
Db 15 GGTACAGGATGGCTGATTGCTGCTGGATCTAATGGCTACTGGAATACAGACATCAG 74
QY 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIle 83
Db 75 GCTGATGATGCGCATGCATACCAAACTTTGAAAAAGGTGGCATTAAGATGAAACATTT 134
QY 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 135 GTGGTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 194
QY 104 IleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyrTrpLysGlu 123
Db 195 ATTAATAGCCAGATGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 254
QY 124 AlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGly 143

```

```

Db 255 GATGTCGACTGTGCACACTTCTTTGCTGTTCTGTTGGAATAAAAGTCAGTTAAAGG 314
QY 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTrpAlaAsp 163
Db 315 GGAAGTGAAGAGTGTGTGACGAGTGGCCCAATATGATCATCTTTATATATATACCGAT 374
QY 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 375 CATGGTGTCCAGGTGTGCTCGGGATGCCAACAAATCCTTTACATGTATGTATATATCTG 434
QY 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 435 ATTGAGGTCTTGAAGAAAAACATGCTTCTGGAACATACAAAGCATGGTATTTATCTT 494
QY 204 GluAlaCysGluSerGlySerMetPheGluGlyLleLeuLysLysAsnLeuAsnIleTyr 223
Db 495 GAAGCATGTAATCTGGAAGTATATTCGAGGAGCTTCTCCCTGAGGGTTTAAATATATAT 554
QY 224 AlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysProGluSerTyr 243
Db 555 GCAACGACAGCATCTGTGTCAGATGAGAACAGTTGGGGTACTTACTGCCCTGGAGATTAT 614
QY 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 615 CCCAGTCCACCTCCAGAAATATGATCCTGTTGGGTGATTGTACAGCTGTTCATGATG 674
QY 264 GluAspSerAspLeuHisAspMetSerLysGluThrLeuGluGlnTyrHisValVal 283
Db 675 GAGACTGTGATGTACACAACTTCGGAACAAACATCAGGAGCAATATCATCTCGTT 734
QY 284 LysArgArgValGlySer 289
Db 735 AAAGAAAGAACCTCAAT 752

RESULT 6
BQ871135
LOCUS      746 bp mRNA linear EST 15-AUG-2002
DEFINITION QG110P16.yg.abl QG_ARCDI lettuce salinas Lactuca sativa cDNA clone
            QG110P16, mRNA sequence.
ACCESSION  BQ871135
VERSION     BQ871135.1 GI:22257685
KEYWORDS   EST.
SOURCE     Lactuca sativa.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
REFERENCE   1 (bases 1 to 746)
AUTHORS    Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevallier,P., Ziegler,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmudson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig7059, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QG110 row: P column: 16.

FEATURES
source       Location/Qualifiers
              1..746
                /organism="Lactuca sativa"
                /cultivar="Salinas"
                /db_xref="taxon:4236"
                /clone="QG110P16"

```

```

/clone_lib="QC-ABCDI lettuce salinas"
/lab_host="E.coli"
/notes="Vector: pBRCDNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG_LIB=QC-ABCDI lettuce salinas
TAG_TISSUE=ROOTS
TAG_SEQ=TTGCCAGGG"

```

```

BASE COUNT 226 a 129 c 187 g 204 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,01e-102 Length: 746
Score: 921.00 Matches: 162
Percent Similarity: 79.42% Conservative: 31
Best Local Similarity: 66.67% Mismatches: 50
Query Match: 37.53% Indels: 0
DB: 14 Gaps: 0

```

```

US-09-934-066-2 (1-466) x B0871135 (1-746)

```

```

Qy 44 GlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTrpAsnTyrArgHisGln 63
Db 15 GGTACACGATGGCTGTATTGCTGCTGATCTAATGCTACTGGAATATACAGACATCAG 74
Qy 64 AlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAspGluAsnIle 83
Db 75 GCTGATGATGCATGCATACCAAACTTTGAAAAAGCTGGCATTAAGGATGAAACATT 134
Qy 84 IleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgProGlyValIle 103
Db 135 GTGGTTTTCATGATGATGACATTCCTCAATAATAGGAGAACCCAGCGCTGGATCATT 194
Qy 104 IleAsnLysProAspGlyLysValPheValPheLysGlyValProLysAspTyrThrLysGlu 123
Db 195 ATTATAGCCAGATGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 254
Qy 124 AlaValAsnValGluAsnValGluAsnValGluAsnValGluAsnValGluAsnValGlu 143
Db 255 GATGAGTGTGGACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 314
Qy 144 GlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAsp 163
Db 315 GGAAGTGGAAAGGTGTGAGCAGTGGCCCAATGATGATGATGATGATGATGATGATGATGAT 374
Qy 164 HisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAlaLysAspPhe 183
Db 375 CATGGTGTCCAGGTGTGCTGGGATGCCAAATCCTTACATGCTATGCTTAATGATCTG 434
Qy 184 AsnGluValLeuGluLysMetHisLysArgLysLysTyrAsnLysMetValIleTyrVal 203
Db 435 ATTGAGGTCTTGAAGAAAAACATCTCTGCAACATACAAAGCATGGTATTTTATCTT 494
Qy 204 GluAlaCysGlySerGlySerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyr 223
Db 495 GAAGCATGTGAATCTGGAAGTATATTCAGGAGCATCTCTCCCTGAGGTTTAAATATATAT 554
Qy 224 AlaValThrAlaAlaAsnSerLysCysSerSerPheValTyrCysProGluSerTyr 243
Db 555 GCAACGACAGCATCTGCTCCAGATGAGAACAGTGGGTACTTACTGCCCTCGGAGATAT 614
Qy 244 ProProProSerGluIleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeu 263
Db 615 CCCATGCCACCTCCAGAAATATGATACCTGTGTGGTGTATTTGATGATGATGATGATGATG 674
Qy 264 GluAspSerAspLeuHisAspMetSerLysLysLysLysLysLysLysLysLysLysLys 283

```

```

Db 675 GAGGACTGTGTATACACAACTCGCAGACAGAAACAATCAGCAGCAATATCATCTGTT 734
Qy 284 LysArgArg 286
Db 735 AAAGAAAGA 743
RESULT 7
LOCUS BG645154 796 bp mRNA linear EST 24-APR-2001
DEFINITION EST506773 KV3 Medicago truncatula cDNA clone pKV3-39K10 5' end,
ACCESSION BG645154
VERSION BG645154.1 GI:13780266
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
Medicago.
REFERENCE 1 (bases 1 to 796)
AUTHORS VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Bowman
,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula 72 h after Rhizobium
inoculation, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbcs.umn.edu
M393561e TIGR sequence name: MTEBX65TK More information is
available at: www.medicago.org
Seq primer: SKmod (CTA GAA CTA gtg gAT CC).
FEATURES
Location/Qualifiers
1..796
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV3-39K10"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
melliotti"
/lab_host="E. coli strain XLOLR"
/notes="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XLOLR cells."
BASE COUNT 229 a 136 c 193 g 238 t
ORIGIN
Alignment Scores:
Pred. No.: 5.92e-101 Length: 796
Score: 907.00 Matches: 161
Percent Similarity: 79.60% Conservative: 38
Best Local Similarity: 64.40% Mismatches: 50
Query Match: 36.96% Indels: 1
DB: 12 Gaps: 0
US-09-934-066-2 (1-466) x BG645154 (1-796)
Qy 38 SerAspLysSerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyr 57
Db 27 AATGATGACAAACGATCAAGGAGCTAGGTGGCTATTTACTTGGTGTCTAATGTTAT 86
Qy 58 TyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGly 77

```

```

Db 87 TGAATATATAGGCATCAGCGTGATGTTGTTCATCCGCTATCAATTTGTTGAGAAAGGTGGC 146
Qy 78 LeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPheSerSerGluAsn 97
Db 147 TTGAAGAAGAAACATTAATTTGTTTCATGATCATGATGATGTTCCCAATGTAGAGAT 206
Qy 98 ProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValPro 117
Db 207 CCAAGGCTGAGTGCATATAATTAACAAACCTGATGCTGTGATGTTTATGAAGGAGTTCCA 266
Qy 118 LysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuGlyAsn 137
Db 267 AAGATTACACTGTCAGAGGTACATGCTGACAAATTTCTATGCTGCTTACTTGGAAAT 326
Qy 138 GluSerGlyValThrGlyAsnGlyLysValLysSerGlyProAsnAspAsnIle 157
Db 327 AATCAGCTCTTACAGGTGGAGTGGGAAAGTTGTGGATAGTGTGCTCCCAATGATCATATT 386
Qy 158 PheIleTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGlu 177
Db 387 TTTGTATACACTGATCATGATGAGGTCAGGGGCTTCTGTGATGTCGCCGTTGGTCTTAC 446
Qy 178 ValMetAlaLysAspPheAsnGluValLeuGlyLysMetHisLysArgLysLysTyrAsn 197
Db 447 TTGTATGCACTGATCTGAATGAAGTCTTGAAGAAACATGCTTCTGGATCATATAAG 506
Qy 198 LysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLys 217
Db 507 AGCCTAGTATTTATCTGGAGGCTGTAATCTCGCAGTATATTTGAAGGACTTCTTCCA 566
Qy 218 LysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyVal 237
Db 567 GAAGATATCAATATCTATGACACACGCTTCAATGCACTAGAGAAAGCAGTTGGGGAACA 626
Qy 238 -TyrCysProGluSerTyrProProProSerGluIleGlyThrCysLeuGlyAspTh 257
Db 627 TAATGCGCTGGGAGTACCTCCCTCCAGAGTACTCAACCTGTAGGTGACCT 686
Qy 257 rPheSerIleSerTrpLeuGluAspSerAspLeuHisAspMetSerLysGluThrLeuG1 277
Db 687 ATACAGTATTGCTGGATGGAAGACAGTGACATACACAATTTACGAAGTGAAGTTTGA 746
Qy 277 uGlnGlnTyrHisValValLysArgArg 286
Db 747 CCCAGCATATAAATTTGTTAAAGATAGG 774

RESULT 8
LOCUS BG646037
DEFINITION EST507656 KV3 Medicago truncatula cDNA clone pkv3-48E6 5' end, mRNA
sequence.
ACCESSION BG646037
VERSION BG646037.1 GI:13781149
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 711)
VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Bowman
,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
ESTs from roots of Medicago truncatula 72 h after Rhizobium
inoculation, 2001
Unpublished (2001)
Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenbos@bcs.umn.edu

```

```

M394444e TIGR sequence name: MTECJ27TK More information is
available at: www.medicago.org
Seq primer: SKmod (CTA gaa cta gta gat CC).
FEATURES
    source
        1..711
            /organism="Medicago truncatula"
            /cultivar="genotype Al7"
            /db_xref="taxon:3880"
            /clone_lib="pkv3-48E6"
            /clone_lib="KV3"
            /tissue_type="Seedling roots"
            /dev_stage="3 days post-inoculation with Sinorhizobium
            meliloti"
            /lab_host="E. coli strain XL0LR"
            /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
            XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
            was directionally ligated into the Unizap XR vector from
            Stratagene and packaged using Gigapack III Gold packaging
            extracts. Plasmids containing cDNA inserts were excised
            from the recombinant lambda-Zap phage using Ex-assist
            helper phage and propagated in XL0LR cells."
BASE COUNT      200 a 121 c 174 g 216 t
ORIGIN
Alignment Scores:
Pred. No.:      1..18e-100      Length:      711
Score:          904.00          Matches:    159
Percent Similarity: 82.28%      Conservative: 36
Best Local Similarity: 67.09%   Mismatches:  41
Query Match:     36.84%        Indels:      1
DB:              12            Gaps:         0
US-09-934-066-2 (1-456) x BG646037 (1-711)
Qy 41 SerAlaLysGlyThrArgTrpAla-ValLeuValAlaGlySerAsnGluTyrTyrAsnTy 60
Db 2 AACGCTCAAGGACTAGTGGGCTATTTACTTGTGGTCTTAATGTTATTGGAATTA 61
Qy 60 rArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAs 80
Db 62 TAGGCATCAGCTCATGTTTGTCTCATGCTATCAATTTGTTGAGAAAGTGGCTTGAAGA 121
Qy 80 pGluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPr 100
Db 122 AGAAACATTAATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
Qy 100 oGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTy 120
Db 182 TGGAGTCAATAATTACAAACCTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 241
Qy 120 rThrLysGluAlaValAlaAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerG1 140
Db 242 CACTGGTGCAGAGGTACATGCTGACAAATTTCTATGCTGCTTTACTTGGAAATAAATCAGC 301
Qy 140 yValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTy 160
Db 302 TCTTACAGGTGGAGTGGGAAAGTTGTGGATGATGATGATGATGATGATGATGATGATGAT 361
Qy 160 rTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGlyAspGluValMetAl 180
Db 362 CTACACTGATCATGGAGTCCAGGGGTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 421
Qy 180 aLysAspPheAsnGluValLeuGlyLysMetHisLysArgLysLysTyrAsnLysMetVa 200
Db 422 ATCTGATCTGAATCAAGTCTTGAAGAAACATGCTTCTGATCATATAAGAGCTTAGT 481
Qy 200 lIleTyrValGluAlaCysGluSerGlySerMetPheGluGlyIleLeuLysLysAsnLe 220
Db 482 ATTTTATCTGGAGCATGCTGAATCTGGCAGTATATTTGAAGGACTTCTTCCAGAGATAT 541
Qy 220 uAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrpGlyValTyrCysPr 240
Db 542 CAATATCTATCGCAGACAGGCTTCAATGATGAGAAAGCAGTTGGGGAACATATTGCCC 601

```


taken from greenhouse plants (4-6wks old TA496). Tissue was immediately frozen in liquid nitrogen."

```

BASE COUNT      216 a 136 c 186 g 228 t
ORIGIN

Alignment Scores:
Pred. No.:      1,28e-99      Length:      766
Score:          896.00      Matches:    162
Percent Similarity: 79.05%      Conservative: 38
Best Local Similarity: 64.03%      Mismatches: 52
Query Match:    36.51%      Indels:    2
DB:            12      Caps:    1

US-09-934-066-2 (1-466) x BGL123960 (1-766)

QY 50 LeuValAlaGlySerAsnGluTyrTyrAsnTyrArgHisGlnAlaAspIleCysHisAla 69
DB 2 CTTCTGCCGGATCAAAATGTTATGGAATTATAGACATCAGGCTGATGTATGCCATCGC 61
QY 70 TyrGlnIleLeuArgLysGlyGlyLeuLysAspGluAsnIleIleValPheMetTyrAsp 89
DB 62 TATCAGCTATTGAGAAAGGTGCTCAAGATGAAATATTATTGCTTCATGTATGAT 121
QY 90 AspIleAlaPheSerSerGluAsnProArgProGlyValIleIleAsnLysProAspGly 109
DB 122 GACATTGCTCACCATCAAGAGAACCCCAAGACCAGAGATTATTATTATAGTCTCGCGGT 181
QY 110 GluAspValTyrLysGlyValProLysAspTyrThrLysGluAlaValAsnValGlnAsn 129
DB 182 GAGATGTTTACGAAGGATCTCTAGGATTACCGGGAGATGATGTTAATGTGCACAC 241
QY 130 PheTyrAsnValLeuLeuGlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValVal 149
DB 242 TTTTGTAGCTGTTCTCTGTGTAACAACAGTCTTACTGAGGTAGCGGAAGGTGGTG 301
QY 150 LysSerGlyProAsnAspAsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeu 169
DB 302 AATAGTGGTCCAATGATCATATTTTCATATTCTATAGTATGATGATGTCGCGCGTG 361
QY 170 IleAlaMetProThrGlyAspGluValMetAlaLysAspPheAsnGluValLeuGlyLys 189
DB 362 CTGGGATGCTACCAATCTTATCTATATGCGGATGATCTAATGCTGTGTTGAAAAG 421
QY 190 MethLysArgLysLysTyrAsnLysMetValIleTyrValGluAlaCysGluSerGly 209
DB 422 AAGCATGCCCTGGGACATATAAAGCTTGATTTGTATGTTGAAGCTTCGAGTCTGGA 481
QY 210 SerMetPheGluGlyIleLeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsn 229
DB 482 AGTATATTGAGGACTTCTTCTTAATGGTCTAAATATTATGCCACACAGCTTCAAT 541
QY 230 SerLysGluSerSerTrpGlyValTyrCysProGluSerTyrProProProSerGlu 249
DB 542 GCTGAGAAGACAGCTGGGAACTATTGCTCGGAGATATCTAGTCTCTCTCTCGAA 601
QY 250 IleGlyThrCysLeuGlyAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHis 269
DB 602 TATCAGACTTGTGCTGGTGATTTGTATGCTGTTTCTCGGATGAGACAGTGAATGCAC 661
QY 270 AspMetSerLysLysThrLeuGluGlnGlnTyrHisValValLysArgArgValGlySer 289
DB 662 AACTTGGGACTGAAATTTAGGACGACGATGATCTTGGTCAAAAGAGAACTGCAATG 721
QY 290 AspVal---ProGluThrSerHisValCysArgPheGly 301
DB 722 GAATACTGCCTAT-GGTTCCCATGTCATGCAATTTGGT 759

RESULT 11
AV559121/c 585 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION AV559121 Arabidopsis thaliana green silicles Columbia Arabidopsis
thaliana cDNA clone SQU11e10F 3', mRNA sequence.
ACCESSION AV559121

```

```

VERSION AV559121.1 GI:8730547
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 585)
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
source
1..585
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQU11e10F"
/tissue_lib="Arabidopsis thaliana green silicles Columbia"
/tissue_type="green silicles"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 160 a 110 c 118 g 197 t
ORIGIN

Alignment Scores:
Pred. No.:      2,02e-99      Length:      585
Score:          893.00      Matches:    171
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    36.39%      Indels:    0
DB:            10      Caps:    0

US-09-934-066-2 (1-466) x AV559121 (1-585)

QY 296 HisValCysArgPheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGly 315
DB 585 CATGTATGCCGTTCCGACAGAGAGATGCTTAAAGATTATCTTCTCTTACATTGGA 526
QY 316 ArgAsnProGluAsnAspAsnPheThrPheThrGluSerPheSerProIleSerAsn 335
DB 525 AGAAATCCTGAAACGATAACTTCACTTTCCACGGAATCTCTTCTCCCAATCTCTAAT 466
QY 336 SerGlyLeuValAsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLys 355
DB 465 TCTGGCTTGTCTCAATCCGCGCGATATTCCTCTGCTATACCTCCAGAGAAAGATTCAAAA 406
QY 356 AlaProMetGlySerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsn 375
DB 405 GCTCCAATGGCATCACTTGAAGCAAAAGAGCTCAGAAAGAAATGCTTGACGAAACAAT 346
QY 376 HisArgLysGlnIleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThr 395
DB 345 CATAGGAACAAATTCGATCAGACATTTACAGACATTTCTGCGGCTTTCAGTTAAACAACC 286
QY 396 AsnValLeuAsnLeuLeuThrSerThrArgThrThrGlyGlnProLeuValAspAspTrp 415
DB 285 AATGCTCTTAATCTCTTAATCTCCACAAAGAACACAGGACGCTCTTTGTAGACGATTGG 226
QY 416 AspCysPheLysThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyr 435
DB 225 GATTGCTTCAAGACTCTAGTTAATAGTTCACGAATCACTGCGGTGCAACGGTGCAATTAC 166
QY 436 GlyLeuLysTyrThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGln 455
DB 165 GGATTGAAGTATACAGGAGCGCTTGCCAATATCTGCAATATGGGAGTGGATGTGAAGCAA 106

```

QY 456 ThrValSerAlaIleGluGlnAlaCysSerMet 466
 Db 105 ACTGTTTCAGCCATTGAACACAGCTTGTTCGATG 73

RESULT 12
 LOCUS BG581593
 DEFINITION EST1483328 GVN Medicago truncatula cDNA clone pGVN-65E16 5' end,
 mRNA sequence.
 ACCESSION BG581593
 VERSION BG581593.1 GI:13596657
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE
 AUTHORS 1 (bases 1 to 776)
 TITLE Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Town
 C.D., Van Aken, S., Utterback, T., Cho, J., and Fraser, C.M.
 ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula, 2001
 JOURNAL Unpublished (2001)
 COMMENT Contact: Carroli P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 University of Minnesota name: M382654e.TIGR sequence name:
 M382654e.TIGR sequence name: M382654e.TIGR sequence name:
 M382654e.TIGR sequence name: M382654e.TIGR sequence name:
 Seq primer: Skmod (CIR 9AA CTA 9TG 9AT CC).

FEATURES
 source
 1..776
 /organism="Medicago truncatula"
 /cultivar="genotype Al7"
 /db_xref="taxon:3880"
 /clone="pGVN-65E16"
 /clone_lib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month
 post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 effective root nodules harvested one month post
 inoculation with Sinorhizobium meliloti. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells."

BASE COUNT 217 a 155 c 186 g 218 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2 8e-96 Length: 776
 Score: 869.00 Matches: 154
 Percent Similarity: 81.36% Conservative: 38
 Best Local Similarity: 65.25% Mismatches: 43
 Query Match: 35.41% Indels: 1
 DB: 12 Gaps: 0

US-09-934-066-2 (1-466) x BG581593 (1-776)

QY 36 GluSerSerAspLysSerAlaLysGlyThrArgTrpAlaValLeuAlaGlySerAsn 55
 Db 69 CAATCCGATGACGATTAACAAATGAAGCCACCAATGGCCATCTTAATGCTGGTCTTAAT 128

QY 56 GluTyrAsnTyrArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLys 75

Db 129 GCCTACTGCAATTACAGACATCAGTCTGATCTTTCTCACGCGTATCAAGTGTTCAGGAAA 188
 QY 76 GlyGlyLeuLysAspGluAsnIleValPheMetTyrAspAspIleAlaPheSerSer 95
 Db 189 GGTGGTTTGAAGAGAGAGACATATTGTTGTTTATGATGACATTTCCAGACANCCAA 248
 QY 96 GluAsnProArgProGlyValIleIleAsnLysProAspGlyGluAspValTyrLysGly 115
 Db 249 GAGAATCCACGCCGCCGAGTCAATTAACAGTCCACATGGAGATGATGTTTACAAAGGA 308
 QY 116 ValProLysAspTyrThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeu 135
 Db 309 GTCCCTTAAGGATTATACCGGTGATGATGTTAATGTTAAACAACCTCTCTTTCCTTACTT 368
 QY 136 GlyAsnGluSerGlyValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAsp 155
 Db 369 GGAATAATAGTCACTCTTACTGCTGCCAGTGGAAAGTTGTGGATAGTGGCTCTTAATGAT 428
 QY 156 AsnIlePheIleTyrTyrAlaAspHisGlyAlaProGlyLeuIleAlaMetProThrGly 175
 Db 429 CACATATTTATATATAGTATGATCAGCGCGTCTCTGGAGTCTGGGATGCTTACTGCT 488
 QY 176 AspGluValMetAlaLysAspPheAsnGluValLeuGluLysMetHisLysArgLysLys 195
 Db 489 CCATTTTATCTACGCGACTGATCTAATTAAGTCTTGAAGAAGACGATCTTCTGAAACT 548
 QY 196 TyrAsnLysMetValIleTyrValGluAlaCysGluSerGlySerMetPheGluClyTle 215
 Db 549 TATAAAGCCTAGTATTTATCTAGAGCATCGAATCTGGAGTATCTTTGACGGTCTT 608
 QY 216 LeuLysLysAsnLeuAsnIleTyrAlaValThrAlaAlaAsnSerLysGluSerSerTrp 235
 Db 609 CTTCCTCAAGGTCTGANTATCTATGACACAGCGCGCAATGCAAGAAGACAGTTCG 668
 QY 236 GlyValTyrCysProGluSerTyrProProProSerGluIleGlyThrCys-LeuG 255
 Db 669 GGAACATATTGCTCTGGGAGAAATCTCTAGTCTCTCCCGCCAGCAATATGAACAACATGCGTTCGG 728
 QY 255 YAspThrPheSerIleSerTrpLeuGluAspSerAspLeuHisAsp 270
 Db 729 TGACCTATACAGTGTGCTTGGATGCAAGACAGTGCATACACAAT 774

AV559478 561 bp mRNA linear EST 07-SEP-2000
 AV559478 Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone S0118b02F 3', mRNA sequence.
 AV559478
 AV559478.1 GI:8730904
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 561)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)
 20363093
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.
 Location/Qualifiers
 1..561
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES
 source

```

/clone="S0118b02F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 154 a 105 c 113 g 189 t
ORIGIN
Alignment Scores:
Pred. No.: 3 08e-96 Length: 561
Score: 867.00 Matches: 167
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.33% Indels: 0
DB: 10 Gaps: 0
US-09-934-066-2 (1-466) x AV559478 (1-561)
QY 300 PheGlyThrGluLysMetLeuLysAspTyrLeuSerSerTyrIleGlyArgAsnProGlu 319
Db 559 TTCGGAACAGAGAAGATGCTTAAGATTATCTTCTTACATTGCAAGAAATCTGAA 500
QY 320 AsnAspAsnPhetThrGluSerPheSerProIleSerAsnSerGlyLeuVal 339
Db 499 AACGATAACTTCACTTTCACGAATCTTTCTCACCACATCTCTAATCTGGCTTGGTC 440
QY 340 AsnProArgAspIleProLeuLeuTyrLeuGlnArgLysIleGlnLysAlaProMetGly 359
Db 439 AATCCGCGGATATCTCTGCTATACCTCCAGAGAAAGATTCAAAAGCTCCAAATGGGA 380
QY 360 SerLeuGluSerLysGluAlaGlnLysLysLeuLeuAspGluLysAsnHisArgLysGln 379
Db 379 TCACCTTGAAGCAAGAAAGCTCAGAGAATGCTTGAGCAAGAAAGATCATAGGAAACAA 320
QY 380 IleAspGlnSerIleThrAspIleLeuArgLeuSerValLysGlnThrAsnValLeuAsn 399
Db 319 ATCGATCAGAGCATACACACATCTCGCGCTTTCAGTTAAACAAACCAATGCTTAAAT 260
QY 400 LeuLeuSerThrArgThrThrGlyGlnProLeuValAspAspTrpAspCysPheLys 419
Db 259 CTCCTTAACCTCCACAAGAAACAACAGACAGACAGCTCTTGAGACGATGGGATTCCTCAAG 200
QY 420 ThrLeuValAsnSerPheLysAsnHisCysGlyAlaThrValHisTyrGlyLeuLysTyr 439
Db 199 ACTCTAGTTAATAGCTTCAAGAAATCACTCGGTGCAACGGTGCAATACGAGATTGAAGTAT 140
QY 440 ThrGlyAlaLeuAlaAsnIleCysAsnMetGlyValAspValLysGlnThrValSerAla 459
Db 139 ACAGGAGCGCTGGCCAAATATCTGCAATATGGGAGTGGATGTGAAGCAAACTGTTTCAGCC 80
QY 460 IleGluGlnAlaCysSerMet 466
Db 79 ATTGAACAAGCTGTGTCGATG 59
RESULT 14
AV555212 538 bp mRNA linear EST 06-SEP-2000
LOCUS
DEFINITION
Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone S0008c10f 3', mRNA sequence.
ACCESSION
AV555212
VERSION
AV555212.1 GI:8726627
KEYWORDS
EST.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 538)
ASANIZU, E., NAKAMURA, Y., SATO, S. and TABATA, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
JOURNAL
MEDLINE 20363093
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Location/Qualifiers
1..538
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="S0008c10f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 168 a 96 c 116 g 158 t
ORIGIN
Alignment Scores:
Pred. No.: 8 76e-95 Length: 538
Score: 855.00 Matches: 163
Percent Similarity: 98.20% Conservative: 1
Best Local Similarity: 97.60% Mismatches: 3
Query Match: 34.84% Indels: 0
DB: 10 Gaps: 0
US-09-934-066-2 (1-466) x AV555212 (1-538)
QY 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuPhe 20
Db 38 ATGCTCTAGTCTCTCTGTCACCTTTCAGATTTCTGGGTGCTCTTTCTTCTTATCTTC 97
QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspAspValGluSerSerAspLys 40
Db 98 TCAGCTGAGTCCCGCAAAACCCCAATTCCTGCAACCATATGATGTTGAATCTACGACAAG 157
QY 41 SerAlaLysGlyThrArgTrpAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
Db 158 AGTGCAAAAGGCACACGATGGGCTGTTTAGTTGCTGGATCAAAATGAATATTAATACTAC 217
QY 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyGlyLeuLysAsp 80
Db 218 AGGCATCAGCGTGACATATGCCACGCTATCAGATATCCGAAAGGCGGCTTTAAAGAT 277
QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerSerGluAsnProArgPro 100
Db 278 GAAACATCATGTGTTTATGATGATATGCGGTTTCTCGGAGAAATCTTAGGCCT 337
QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
Db 338 GGAGTTATCATTAATAACAGATGAGAGATGTTTATAAAGGAGTTCCTAAGGACTAC 397
QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
Db 398 ACTAAAGAGCTGTTAATGTTCAAAACTTCTACAATGTGTTACTTGAATGAAGCGGC 457
QY 141 ValThrGlyGlyAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
Db 458 GTCACAGAGGAAGAAATGGCAAGTTGTGAAAAGTGGCCCTAATGATATATCTTCATCTAT 517
QY 161 TyrAlaAspHisGlyAlaPro 167
Db 518 TATGCTGACCATGGAGCTCCT 538
RESULT 15
AV562667 535 bp mRNA linear EST 07-SEP-2000
LOCUS
DEFINITION
Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone S0173h03f 3', mRNA sequence.
ACCESSION
AV562667
VERSION
AV562667.1 GI:8734093
KEYWORDS
EST.

```

SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Euclidytedons; Core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 535)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source
 1..535
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ173h03F"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 170 a 92 c 114 g 159 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,36e-94 Length: 535
 Score: 848.00 Matches: 162
 Percent Similarity: 98.79% Conservative: 1
 Best Local Similarity: 98.18% Mismatches: 2
 Query Match: 34.56% Indels: 0
 DB: 10 Gaps: 0
 US-09-934-066-2 (1-466) x AV562667 (1-535)
 QY 1 MetSerSerProLeuGlyHisPheGlnIleLeuValPheLeuHisAlaLeuLeuIlePhe 20
 DB 40 ATGTTTAGTCTCTTGGTCACCTTCAGATCTCTGTCTTCTTATGCTTTGCTTATCTTC 99
 QY 21 SerAlaGluSerArgLysThrGlnLeuLeuAsnAspValGluSerSerAspLys 40
 DB 100 TCAGCTGAGTCCGCAAAACCAATTCGTAACGATATGATGTTGAATCTAGCGACAAG 159
 QY 41 SerAlaLysGlyThrArgTIPAlaValLeuValAlaGlySerAsnGluTyrTyrAsnTyr 60
 DB 160 ACTGCAAAAGGCACACGATGGCTGTTTACTGCTGATCAATGAATATTATAACTAC 219
 QY 61 ArgHisGlnAlaAspIleCysHisAlaTyrGlnIleLeuArgLysGlyLeuLysAsp 80
 DB 220 AGGCATCAGGTGACATATGCCCGGTATCAGATCTCCGAAAGCGGTTTAAAGAT 279
 QY 81 GluAsnIleIleValPheMetTyrAspAspIleAlaPheSerGluAsnProArgPro 100
 DB 280 GAAACATCATGTGTTTATGATGATGATATCGGTTTCTCGAGAAATCTTAGGCCT 339
 QY 101 GlyValIleIleAsnLysProAspGlyGluAspValTyrLysGlyValProLysAspTyr 120
 DB 340 GGAGTTATCATTAATAAACACGATGGAGAGATGTTTATAAAGGAGTTCCTAAGGACTAC 399
 QY 121 ThrLysGluAlaValAsnValGlnAsnPheTyrAsnValLeuLeuGlyAsnGluSerGly 140
 DB 400 ACTAAAGAGGTGTTAATGTTTCAAAACCTTACAAATGCTGTACTTGGAAATGAAAGTGC 459
 QY 141 ValThrGlyLysAsnGlyLysValValLysSerGlyProAsnAspAsnIlePheIleTyr 160
 DB 460 GTCACAGGAGGAAATGGCAAAAGTTGTGAAAAGTGGTCTCTATGATAATATCTTCATCTAA 519
 QY 161 TyrAlaAspHisGly 165

Db 520 TATGCTGACCATGGA 534
 Search completed: May 29, 2003, 19:45:27
 Job time : 1612 secs